A MATRIX POLYNOMIAL APPROACH TO QUASI-BIRTH-DEATH PROCESSES AND ITS IMPLICATIONS ON MODELS OF PRODUCTION LINES

By

MEHMET MURAT FADILOGLU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1998

To my family, to my friends, to love

ACKNOWLEDGMENTS

I would like to use this opportunity to express my sincerest appreciation to Dr. Sencer Yeralan for his continuous support throughout this endeavor. His trust in me and his friendship made it possible.

I would also like to convey my gratitude to my committee members, Dr. Eginhard J. Muth, Dr. Suleyman Tufekci, Dr. Sherman X. Bai, and Dr. John K. Schueller.

Finally, I would like to mention the names of those people whose moral support was indispensable for me all the way to this point: Cicek Fadiloglu, Ahmet Fadiloglu, Tolga Fadiloglu, Gamze Fadiloglu, Fahriye Yigiterhan, Gonul Yigiterhan, Elcin Yildirim, M. Emin Alpay, Utku Unver, Vanesa C. Anthony.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS iii

ABSTRACT vi

INTRODUCTION 1
1.1 Topic Area, Extent and Motivation 1
1.2 Contribution and Its Significance 3

PAST WORK 5

BACKGROUND AND MODELS 10
3.1 Introduction 10
3.2 Introduction to Quasi-Birth-Death Processes: 11
3.3 QBDP Models for Production Lines 15
Model 1: Exponential Server with Limited Buffer 16
Model 2: Exponential Server with Breakdown and Repair, Limited Buffer 18
Model 3: Server With Erlang-2 Service Times, Limited Buffer 20
Model 4: Server With Erlang-2 Service Times, Breakdown and Repair, Limited Buffer 23
Model 5: Two Exponential Servers with One Intermediate Limited Buffer 26
Model 6: Two Exponential Servers With One Intermediate Limited Buffer, Breakdown and Repair 28
Model 7: Server With Erlang-N Service Times, Limited Buffer 33
Model 8: Two Exponential Servers with Two Limited Buffers 37

A GENERAL THEORY ON SPECTRAL PROPERTIES OF STATE-HOMOGENEOUS FINITE-STATE QUASI-BIRTH-DEATH PROCESSES 43
4.1 Introduction 43
4.2 The Matrix Polynomial Solution Procedure 45
4.3 Spectral Properties of State-Homogeneous Finite-State Quasi-Birth-Death Processes 51

IMPLICATIONS OF THE SPECTRAL THEORY ON MODELS OF PRODUCTION LINES 73
5.1 Introduction 73
5.2 General Implications 74

5.3 Application of the Spectral Theory to Specific Models of Production Lines 77
Model 1: Exponential Server With Limited Buffer: 78
Model 2: Exponential Server with Breakdown and Repair, Limited Buffer: 81
Model 3: Server With Erlang-2 Service Times, Limited Buffer: 99
Model 4: Server With Erlang-2 Service Times, Breakdown and Repair, Limited Buffer: 105
Model 5: Two Exponential Servers With One Intermediate Limited Buffer 108
5.4 Common Spectral Characteristics of Models of Production Lines 116

CONCLUSION 120

LIST OF REFERENCES 125

BIOGRAPHICAL SKETCH 128

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A MATRIX POLYNOMIAL APPROACH TO
QUASI-BIRTH-DEATH PROCESSES
AND ITS IMPLICATIONS
ON MODELS OF PRODUCTION LINES

By

Mehmet Murat Fadiloglu

August 1998

Chairman: Sencer Yeralan
Major Department: Industrial and Systems Engineering

Birth-death processes are quite pervasive in the domain of stochastic modeling. They are defined on an ordered discrete set that is called the counting set. A birth-death process can make a transition to a neighboring state–an element of the counting set–at any given time instant or remain at the same state. The transition times (time to next such transition) are exponentially distributed. Thus, they are categorized as Markov chains and they are amenable to all solution procedures for Markov chains.

The quasi-birth-death process is a generalization of the same idea. Instead of having one state corresponding to an element of the counting space, one has a group of

states corresponding to an element. Consequently, the transition rate matrix, which has a tridiagonal structure in the birth-death processes, manifests a nested block tridiagonal structure.

We present a series of production systems models and depict how the models fit into the framework of quasi-birth-death processes. Thus, we demonstrate the power of the processes from a modeling perspective.

Consequently, we apply a matrix polynomial method to obtain the steady-state probabilities in state-homogeneous finite-state quasi-birth-death processes. The method is based on finding the eigenvalue-eigenvector pairs that solve a matrix polynomial equation. Since the computational effort in the solution procedure is independent of the cardinality of the counting set, it has an immediate advantage over other solution procedures.

We present and prove different properties relating the quantities that arise in the solution procedure. By also compiling and formalizing the previously known properties, we present a formal unified theory on the spectral properties of quasi-birth death processes, which furnishes a formal framework to embody much of the previous work. This framework furthers our understanding of the behavior the modeled systems manifest.

Finally, by applying this framework to the production line models we depict, we demonstrate how the spectral theory we have discovered can be applied to specific instances of quasi-birth processes. Furthermore, we gather new results that contribute to stochastic analysis of production line models. By contributing to this well-established area of research, we also evince the value of the spectral theory we derived.

In retrospect, we developed a theoretical tool that can be used in the analysis and design of production lines, which form the backbone of manufacturing systems. The importance of contributing new tools to this crucial area lies in sustaining of the economic competitiveness, which is critical for survival of today's manufacturing enterprises.

CHAPTER 1
INTRODUCTION

1.1 Topic Area, Extent and Motivation

This dissertation aims to provide a formal analysis for the class of stochastic models, which can be categorized under the name of state-homogeneous finite-state Quasi-Birth-Death Processes (QBDP). Although the nature of the work is almost purely mathematical, it is actually an attempt to understand the inner-workings of the QBDPs, which have been extensively used in the domain of stochastic modeling in order to understand the behavior of certain physical systems such as production lines.

QBDPs are stochastic processes, more specifically Markov processes. As it is in the large domain of stochastic processes, one would like to statically characterize the time evolution of the processes. The procedures to do this kind of analysis are well established for Markov processes. Yet, since these analysis procedures are fully numerical, their direct application to QBDPs does not provide any understanding of the system. It only confers numerical values for a specific realization. Moreover, the procedure does not take advantage of the structure manifested by the QBDPs.

For determining the steady-state behavior of the QBDPs, we propose the use of a matrix polynomial technique in order to overcome the deficiencies of the existing procedures for the ordinary Markov processes. That was first proposed by Tan (1994). In their seminal work, Tan and Yeralan investigated the basic properties of relating the

quantities that arise in the solution procedure. In this dissertation, we extend this work by providing new properties that bestow further insights about the QBDPs. Our objective is to unite all the properties within a formal framework.

The initial motivation behind all the analyses that will be presented within this dissertation was to get a better understanding of production lines and to develop techniques that could be used in their analysis. The basic models of production lines that encompass one buffer all give rise to QBDPs (Yeralan and Muth 1987). Indeed, all production systems with buffers give rise to QBDPs. In this sense, one of the buffer levels corresponds to the counting set of the process. When more buffers are present, the state-space that corresponds to each element of the counting set becomes much larger. The proposed solution technique's computational effort is independent of the size of the counting space. Using the technique, we can even provide a closed-form solution for different counting set cardinalities. Furthermore, the quantities that manifest themselves in the solution procedure seem to characterize certain properties of the modeled production systems.

The second part of this work will be the investigation of the relation between the theoretical properties that have been formally established and the actual behavior of the systems modeled. It will closely relate to the aforementioned production systems, since the production systems will constitute a concrete platform for the investigation. Although the results obtained will be extendible to a number of other systems that can be modeled with QBDPs, the production systems, which we have been extensively studied, will be a point in case. Our familiarity with these systems will facilitate the investigation.

The properties that relate the quantities arising in the solution procedure grant us clues about the behavior of the modeled system. For instance, we already know that the eigenvalue at one relates to the balance of the systems. If the systems' feeding and output retrieval rates are the same, the multiplicity of the eigenvalue at one is at least two. Yet, this relation is not obvious and needs attentive examination to be fully exposed. The second part, which is based on the findings of the first part, will be of an empirical nature, geared towards the exploitation of the theoretical results for engineering insights and practical use. Different instances of the production systems with known characteristics will be analyzed and an attempt to link the characteristics to the elements of the solution procedure will be made.

1.2 Contribution and Its Significance

Many researchers have worked on the analysis of the production systems and specifically production lines. At the beginning, most such efforts were directed towards analyzing a specific system under a rather limiting set of assumptions. Under these tight provisions, they were able to investigate certain characteristics of the systems. Consequently, the models lacked the generality that would enable us to capture the general characteristics of such systems and to develop analysis procedures that are valid for a large class of actual systems.

On the other hand, simulation was always a viable tool in analyzing production systems. It is possible to incorporate any feature of a physical system into a simulation model. Yet all that simulation provides are numerical values that are valid for an instant

of a model. The typical solution procedures, which are based on generating a Markov model for a system and numerically solving it, also suffer from the same ailment, although the computational effort involved is significantly lower.

Yeralan and Muth (1987) showed that a rather large class of production lines with different properties could be modeled within the framework provided by QBDPs. Subsequently, Tan and Yeralan used the matrix polynomial technique in the analysis of the production lines. They also investigated different properties relating the quantities arising in the solution procedures. This framework is extremely successful in capturing the structure present in many production line models and providing quantities that can be related to physical properties of the system.

This work is another step on the rewarding direction that aforementioned research has established. We provide new properties relating the quantities of interest in the solution procedure. Most importantly, we unify the new and the previously stated properties in a formal theory. Certain relations that were previously stated were not rigorously proven due to the lack of a formal framework. Here, we contribute rigorous proofs to the entire known and discovered properties of the state-homogeneous QBDPs. Thus, we present a formal framework to embody all the previous work on the subject. Furthermore, this framework carries the prospect of furthering the understanding of physical properties of the modeled system.

CHAPTER 2
PAST WORK

This work is strongly connected to two different streams of past research. The first stream is the literature that deals with mathematical analysis of QBDPs. The second stream is the literature that deals with modeling of production lines. The connection of this work with its first stream is more obvious, because of the mathematical formalism with which it attacks the state-homogenous QBDPs. Yet, since the motivation behind all this formal work is to get a better understanding of the production systems, it is in many ways a beneficiary of the past research in the area of production research.

The study of QBDPs was initiated by R. V. Evans (1967) and the Ph.D. thesis of V. Wallace (1969). Wallace was the one to coin the term *quasi-birth-death processes*. The most detailed discussion of QBDPs is done by Neuts (1981). In his book, Neuts studies Markov chains using matrix geometric invariant vectors. His methodology is based on the fact that subvectors of the steady-state probability vector are related to one another in a matrix geometric fashion. Neuts also provides a probabilistic interpretation for the rate matrix associated with matrix geometric invariant vectors.

There is a difference between the QBDPs that Neuts attacks and the QBDPs investigated within this work. Neuts' QBDPs are not finite-state, that means the recursive equation that defines the QBDP is repeated infinitely, and there is no upper boundary for the process. Whereas, in the QBDPs of our interest, the repetition of the recursive equation is finite. Moreover, there is an upper boundary equation, which does

not exist in Neuts' case. Neuts' methodology is extended to QBDPs with upper boundaries by Hajek (1982).

The production line literature, being the second stream to which our study is connected, is quite extensive. It begins with the pioneering work on production lines by Koeningberg (1959) and it reaches today with an ever-increasing rate of contribution. The contributions of Muth (1979a) and Buzacott (1967) deserve special credit, not because of the bulk of research they have produced, but because of their influence in advancing the frontier in the field. The excessive nature of related literature forces us to refer the reader to the review paper by Dallery and Gershwin (1992), which is praiseworthy in its thoroughness. This paper includes both discrete and continuous–in terms of time and parts that are produced–models of the production systems and all analytical results obtained in the past. It also includes approximate methods for the analysis of production systems. But since our research methodology is exact, they do not form a direct antecedent for our line of research.

Most of the studies in the field of production line research use assumptions that enables them to utilize the Markov chain methodology. For the models not built under these rather restrictive assumptions, the analysis tools available are extremely scarce and existing tools are difficult to use because of the extensive need for statistical characterization. Most common assumptions in the literature can be cited as

1. Station service times, station breakdown times, and station repair times are statistically independent and all process parameters are time invariant.
2. Station service times, station breakdown times, and station repair times are geometrically or exponentially distributed or have phase type distributions.

3. There are no pauses or delays in the services. That is, the production starts right after the repair is completed and the repair starts right after the breakdown occurs.

An observation that can be made about the bulk of the literature is that almost all of the studies deal with specific models. These works stop short of providing a general framework that can be applied for a large class of production systems. They make use of the special structure of the modeled systems. A handful of papers can be rightfully cited here for they epitomize the aforementioned tendency.

Hunt (1956), Hatcher (1969), Avi-Itzak and Yadin (1965), Rao (1975) presented models of two- and three-station production lines with no station breakdown. Hillier and Boling (1967) examined the effect of line unbalance on the production lines with exponential and Erlang service times and identical buffers. For three-station production lines, they showed that placing the slower station in the middle improves the production rate. This propitious ordering is explained by the so-called "bowl phenomenon." In this research effort we also examine the balance of the systems and show that the balance gives rise to an additional eigenvalue at one within the framework of the matrix polynomial methodology we are proposing.

A similar research history can be observed for the unreliable production lines. Vladzievskii's paper (1952) is most probably the earliest work on production lines subject to breakdown. Among other works that continued this line of research by giving analytical solutions to models of two-station production lines subject to failure, Sevast'Yanov (1962), Koeningsberg (1959), Buzacott (1967), Sheskin (1974), Okamura and Yamashina (1977), Schick and Gershwin (1978), Gershwin and Berman (1981),

Berman (1982), and Sastry (1985) are the most prominent ones. Muth's work (1979b) is of distinguished importance, since he expounded the general properties of unreliable production lines without making specific assumptions on service time distributions.

Although this work relates to all the aforementioned literature, it is the immediate successor of an established line of research. The first work in this line is by Yeralan and Muth (1987). Yeralan and Muth show that the QBDPs are a viable tool for modeling of a large class of production lines. They give eight models, which differ from each other in the assumptions concerning station breakdown, station repair and transfer mechanisms within the framework of QBDPs. They propose three different solution procedures to be used in conjunction with the modeling framework. The first one is the general recursive solution procedure, which works for all possible models within the framework. The others are the matrix geometric and scalar geometric procedures that can be applied to models that have some additional structure. They also relate their methodology to Neuts' work.

The second in the aforementioned line research is the work of Yeralan and Tan (1997). This work is the first to propose the application of matrix polynomial techniques in order to find the steady-state probabilities of the QBDPs. Subsequently, the techniques are applied to the models of production lines. In a subsequent paper (Tan and Yeralan 1997), they use the findings of the aforementioned work as a building block for a decomposition method that tackles production lines with many stations.

Furthermore, the methodology proposed by Yeralan and Tan also gives certain clues that can be of use in understanding the properties of the modeled physical systems. The relations between the physics of the system and the quantities that arise in the

solution procedure are shown by a series of properties that delineate the fact that certain elements of the solution procedure exhibit certain physically meaningful behavior. Thus, one should be able to use these quantities in order to tell certain characteristics of the physical systems.

This body of the work extends the research of Tan and Yeralan (1997) by providing new properties that bestow further insights about the QBDPs and then unites all the properties within a formal framework. Then by applying this framework on the representative models of production lines, it contributes to the long progression of the frontier in the area of production line research.

CHAPTER 3
BACKGROUND AND MODELS

3.1 Introduction

The aim of this chapter is to present the groundwork needed to put the rest of this research effort on a meaningful setting. With this end in mind, we first make an introduction to QBDPs, so that the reader who may not be acquainted with this aspect of stochastic modeling may introduce himself to the basic foundations of the topic. This is indispensable in order to comprehend further developments presented within this body.

As a part of the effort of elucidating the topic, first, we introduce a formal definition of QBDP and contrast the QBDP with the traditional birth-death processes. Later in the chapter, we present examples of QBDP models that pertain to production lines. Our aim in doing that is to show how QBDPs present themselves within this context and to show the kind of work that needs to be done to specify fully the QBDPs that corresponds to a class of models. By compiling the aforementioned models, we demonstrate how the theory that will be presented can be related to engineering concerns. Eventually, in Chapter 5, these models will be used to illustrate how the theoretical findings in our research can be applied to gain insight about systems that can be modeled by QBDPs.

3.2 Introduction to Quasi-Birth-Death Processes:

The study of QBDPs goes back to R. V. Evans (1967) and to V. Wallace (1969). They are the ones who initiated the field of study. In fact, Wallace coined the term *quasi-birth-death process*. Yet, the most comprehensive study in the field is by Neuts (1981). Since all the following research refers to Neuts for the basics of the field, we find it appropriate to give the definition stated by Neuts:

A *quasi-birth-death process* is a Markov process on the state space $E = \{(i, j), i \geq 0, 1 \leq j \leq m\}$, with the infinitesimal generator $\tilde{Q}$, given by

$$\tilde{Q} = \begin{bmatrix} B_0 & C & & & & & \\ A_0 & B & C & & & & \\ & A & B & C & & & \\ & & A & B & C & \cdots & \\ & & & A & B & \cdots & \\ & & & \cdot & \cdot & & \\ & & & \cdot & \cdot & & \\ & & & \cdot & \cdot & & \end{bmatrix} \tag{3.1}$$

where $B_0\underline{1}^T + C\underline{1}^T = A_0\underline{1}^T + B\underline{1}^T + C\underline{1}^T = A\underline{1}^T + B\underline{1}^T + C\underline{1}^T = \underline{0}$.

The QBDP, as defined, is a generalization of the birth-death process. Birth-death processes are stochastic processes of Markov type where each state can be associated with an element of a counting set, which effectively means that the states can be ordered in a linear fashion. Once the correct ordering is established, transitions should occur only to neighboring states: that means given that the process is in state n, it can have a transition only to state n+1 or state n-1. And the transition times are exponentially distributed

which allows us to classify them under the general category of Markov processes. As with any Markov process, a state transition diagram can be used to represent it.

Figure 3.1 Basic birth-death process model

In the case of QBDP, instead of having one state corresponding to an element of the counting set, we have a group of states corresponding to it. Thus from a given group of states, the process can have transitions to the group of states corresponding to the previous element of the counting set or to the group of states corresponding to the next element of the counting set. This kind of Markov process yields a state transition diagram of the form shown in Figure 3.2.

Figure 3.2 Representation of a QBDP (The dashed lines represent a collection of transitions originating from a state in one group and terminate in a state in another group.)

Yet the framework that we have defined here for a QBDP does not fully comply with the definition of Neuts that we have stated. In Neuts' definition a QBDP is defined by a generator matrix that has a block nested tridiagonal structure in which the same blocks are repeated except for the boundary states. For this to occur, one would need to have the same number of states corresponding to each counting set element. Moreover, the same kind of transitions with the same transition rates should be present in between these states. Thus all literature that adopts the definition of Neuts really relates to state-homogeneous quasi-birth-death processes. Although it is more correct to use the state-homogeneous term for the kind of QBDPs that we are going to delve into, within this endeavor, we are going to omit the "state-homogeneous" part and use the term QBDP as it is used in the literature.

Within the framework defined for QBDPs, it is most appropriate to define the state space as the Kroenicker product of two sets, one of which is the counting set. Thus, the other set is used to distinguish all possible things that can happen at a given counting set level. The same things should be possible at all counting set levels and the way the process goes from one of them to another should be independent of the level.

The QBDPs that Neuts investigates differ in one way from the ones that are the subject of our study. The counting set is infinite, in other words the counting set is the set of natural numbers, for Neuts' case. The methodology that we are going to represent here is for the case of a finite counting space. This means that instead of having only one boundary we obtain two boundaries for our QBDPs.

After describing the kind of QBDPs that we concentrate on, it is appropriate to give a formal definition for them:

A *quasi-birth-death process* is a Markov process on the state space $E = \{(i, j), 0 \le i \le M, 1 \le j \le n\}$, with the transition rate matrix–or the infinitesimal generator– R, given by

$$R = \begin{bmatrix} B_0 & C & & & & & & \\ A & B & C & & & & & \\ & A & B & C & & & & \\ & & \cdot & \cdot & \cdot & & & \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \\ & & & & & A & B & C \\ & & & & & & A & B_M \end{bmatrix} \tag{3.2}$$

where $B_0 \underline{1}^T + C\underline{1}^T = A\underline{1}^T + B\underline{1}^T + C\underline{1}^T = A\underline{1}^T + B_M \underline{1}^T = \underline{0}$.

From this point on, whenever we use the term QBDP, we mean the process defined here.

One can develop a numerical solution procedure for the case of infinite counting set, within the framework of matrix polynomial methods that we are proposing in this study. The solution procedure is actually more general when the counting space is assumed to be finite. For example if the system has modes that grows with the numbers in the counting set (that occurs if there are eigenvalues with absolute values greater than one), the steady-state is never reached for the infinite case, yet if there is an upper boundary, the steady-state is always reached. Thus, our methodology is applicable for those systems that have a tendency to grow with the elements of the counting set. Since the number of eigenvalues increases with the number of states corresponding to each element of the counting space, one can observe that the restriction that is put on the

procedure by forcing all the eigenvalues of the system to be in the unit circle–in the complex numbers plane–is tremendous.

Moreover, our procedure can also accommodate a solution for Markov models that do not yield a strictly block tridiagonal structure in their transition matrices. The boundary equations do not have to relate only to the neighboring state groups, as one would expect. This condition can be relaxed with the framework that is provided.

3.3 QBDP Models for Production Lines

Although the QBDP platform is quite general and can be used for many stochastic models that pertain to different areas, here we are going to give examples of QBDP models that pertain to production lines. Eventually, the theory that we have developed for QBDPs will be applied to these models in order to obtain certain insights. Thus, the production line models will form the applied part of this research effort. A similar effort can be conducted for other areas, which give rise to QBDPs, but we are going to limit the scope of the present research to the models of production lines.

In this section we are going to present the production line models in a natural progression. This means that we are going to present first basic models and then go towards more complex models that mimic further behaviors of real production systems. The submatrices that define the QBDP for each model will be specified. Once these are produced the models are ready for application of the matrix polynomial procedure we are proposing.

Model 1: Exponential Server with Limited Buffer

This is most probably the first model to be encountered in any queuing theory book, the quintessential M/M/1 model. Here the only deviation from this basic model is that the buffer–or the queue–has a limited size, M. In most literature this is named M/M/1/M, with the last M representing the size of the queue. In this model, parts arrive to the server as a Poisson process with a mean rate of λ, and they are serviced with exponential service times with mean μ.

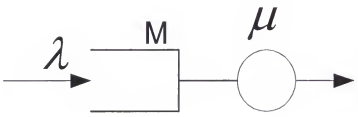

Figure 3.3 Exponential server with limited buffer and its relevant parameters

One can obtain the Markov chain diagram shown in Figure 3.4 using this description for the model. As one can observe from the diagram, this model is actually a simple birth-death process. But since QBDP is a generalization of birth-death process this model still qualifies for the name QBDP. Actually it forms one of the simplest QBDPs possible. One can easily write the steady-state equations for this model by looking at the diagram:

$$-\lambda P(0) + \mu P(1) = 0$$

$$-(\lambda + \mu) P(i) + \mu P(i+1) + \lambda P(i-1) = 0 \quad for\ 1 \le i \le M-1$$

$$-\mu P(M) + \lambda P(M-1) = 0$$

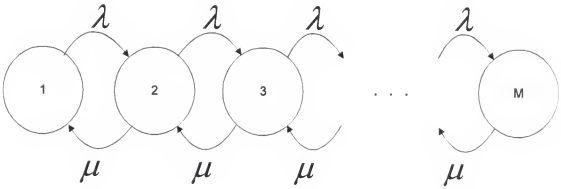

Figure 3.4 Markov transition diagram of Model 1

One can obtain the following transition rate matrix from the preceding equations:

$$R = \begin{bmatrix} -\lambda & \lambda & & & & & & \\ \mu & -(\lambda+\mu) & \lambda & & & & & \\ & \mu & -(\lambda+\mu) & . & & & & \\ & & \mu & . & . & & & \\ & & & . & . & . & & \\ & & & & . & . & \lambda & \\ & & & & & . & -(\lambda+\mu) & \lambda \\ & & & & & & \mu & -\mu \end{bmatrix}$$

Clearly, we have a transition rate matrix that fits the platform defined by our definition of a QBDP. By comparing this transition rate matrix with the generic transition rate matrix defined by equation (3.2), one can give the specific values of the submatrices used in the formal definition:

$$A = \mu \ , \ B = -(\lambda + \mu) \ , \ C = \lambda$$

$$B_0 = -\lambda \ , \ B_M = \ -\mu$$

It is easy to observe that for this model the submatrices in the definition of the QBDP manifest themselves as simple scalars. Although the application of the matrix

polynomial method for this case is overkill, it may help us to understand basic mechanisms of the procedure.

Model 2: Exponential Server with Breakdown and Repair, Limited Buffer

This model builds on Model 1 by incorporating a breakdown and repair. If the server is operating, it breaksdown with a time to breakdown distribution that is exponential with parameter α. Once the breakdown occurs, the repair starts and time to the end of repair is also exponentially distributed with parameter β. During the breakdown, the part that was being processed at the time breakdown occurred is not processed. Processing starts on the same part after the repair.

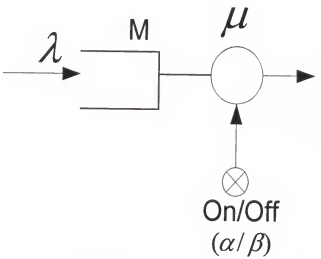

Figure 3.5 Exponential server with breakdown and repair and its relevant parameters.

For the described process the Markov transition diagram is given in Figure 3.6. We can clearly see that the state space is the Kroenicker product of the counting set with a set of two elements {W,R}. W corresponds to the states for which the server is working and R corresponds to the states for which the station is broken and the repair is taking place.

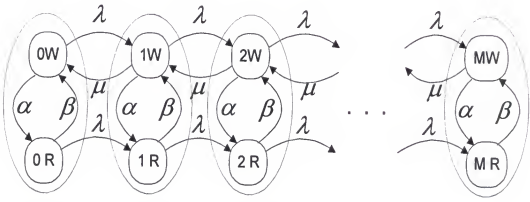

Figure 3.6 Markov transition diagram of Model 2.

The steady-state equations for the process are

$$-(\alpha+\lambda)P(0,W)+\beta P(0,R)+\mu P(1,W)=0$$

$$-(\beta+\lambda)P(0,R)+\alpha P(0,W)=0$$

$$-(\alpha+\lambda+\mu)P(i,W)+\beta P(i,R)+\lambda P(i-1,W)+\mu P(i+1,W)=0 \qquad for\ 1\le i\le M-1$$

$$-(\lambda+\mu)P(i,R)+\alpha P(i,W)+\lambda P(i-1,R)=0 \qquad for\ 1\le i\le M-1$$

$$-(\alpha+\mu)P(M,W)+\beta P(M,R)+\lambda P(M-1,W)=0$$

$$-\beta P(M,R)+\alpha P(M,W)+\lambda P(M-1,R)=0$$

One can obtain the following transition rate matrix from the preceding equations:

$$R = \begin{bmatrix} -(\alpha+\lambda) & \alpha & \lambda & 0 & & & & & & & \\ \beta & -(\beta+\lambda) & 0 & \lambda & & & & & & & \\ \mu & 0 & -(\alpha+\lambda+\mu) & \alpha & \cdots & & & & & & \\ 0 & 0 & \beta & -(\beta+\lambda) & \cdots & & & & & & \\ & & & & \cdots & & & & & & \\ & & & & \cdots & & & & & & \\ & & & & \cdots & & & & & & \\ & & & & \cdots & & & & & & \\ & & & & & \cdots & -(\alpha+\lambda+\mu) & \alpha & \lambda & 0 \\ & & & & & \cdots & \beta & -(\beta+\lambda) & 0 & \lambda \\ & & & & & & \mu & 0 & -(\alpha+\mu) & \alpha \\ & & & & & & 0 & 0 & \beta & -\beta \end{bmatrix}$$

Clearly, this transition rate matrix conforms to the definition of a QBDP that we are employing. The dimension of the submatrices is 2x2 and the submatrices can be easily extracted:

$$A = \begin{bmatrix} \mu & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -(\alpha+\lambda+\mu) & \alpha \\ \beta & -(\beta+\lambda) \end{bmatrix}, \quad C = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$B_0 = \begin{bmatrix} -(\alpha+\lambda) & \alpha \\ \beta & -(\beta+\lambda) \end{bmatrix}, \quad B_M = \begin{bmatrix} -(\alpha+\mu) & \alpha \\ \beta & -\beta \end{bmatrix}$$

Model 3: Server With Erlang-2 Service Times, Limited Buffer

The only difference between this model and Model 1 is the statistical distribution of the service times. In the M/M/1 model the distribution is exponential, whereas here it is Erlang which is the sum of two exponential distributions. In order to accommodate for this non-exponential distribution one needs to incorporate an additional state for each counting level. When this idea is applied, one gets the Markov transition diagram shown in Figure 3.8.

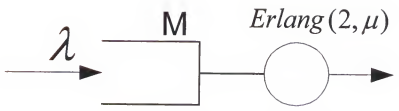

Figure 3.7 Server with Erlang-2 service times, limited buffer and its relevant parameters.

The state space is the Kroenicker product of the counting set and the set {1,2}, which keeps track of the Erlang stage in the processing of parts. Here, only one state corresponds to the element 0 of the counting set. Consequently, this element cannot be incorporated in our QBDP and has to be handled separately. The QBDP starts from element 1. This has to be taken into account while applying the matrix polynomial solution procedure.

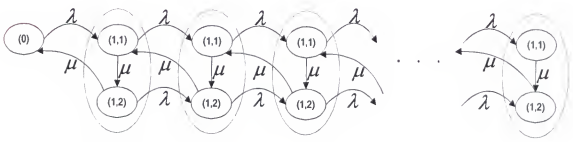

Figure 3.8 Markov transition diagram of Model 3

The steady-state equations for the process are

$$-\lambda P(0) + \mu P(1,2) = 0$$

$$-(\lambda + \mu) P(1,1) + \lambda P(0) + \mu P(2,2) = 0$$

$$-(\lambda + \mu) P(1,1) + \mu P(1,2) = 0$$

$$-(\lambda+\mu)P(i,1)+\lambda P(i-1,1)+\mu P(i+1,2)=0 \qquad for\ 2\le i\le M-1$$

$$-(\lambda+\mu)P(i,2)+\lambda P(i-1,2)+\mu P(i,1)=0 \qquad for\ 2\le i\le M-1$$

$$-\mu P(M,1)+\lambda P(M-1,1)=0$$

$$-\mu P(M,2)+\lambda P(M-1,2)+\mu P(M,1)=0$$

After substituting the value of $P(0)$ obtained from the first equation into the second equation, one can obtain the following transition rate matrix from the preceding equations:

$$R=\begin{bmatrix}
-(\lambda+\mu) & \mu & \lambda & 0 & & & & & & & \\
\mu & -(\beta+\lambda) & 0 & \lambda & & & & & & & \\
0 & 0 & -(\lambda+\mu) & \mu & \cdot & \cdot & & & & & \\
\mu & 0 & 0 & -(\lambda+\mu) & \cdot & \cdot & & & & & \\
 & & & & \cdot & \cdot & \cdot & \cdot & & & \\
 & & & & \cdot & \cdot & \cdot & \cdot & & & \\
 & & & & \cdot & \cdot & \cdot & \cdot & & & \\
 & & & & & \cdot & \cdot & \cdot & -(\lambda+\mu) & \mu & \lambda & 0 \\
 & & & & & \cdot & \cdot & \cdot & 0 & -(\lambda+\mu) & 0 & \lambda \\
 & & & & & & & & 0 & 0 & -\mu & \mu \\
 & & & & & & & & \mu & 0 & 0 & -\mu
\end{bmatrix}$$

Clearly this transition rate matrix conforms to the definition of a QBDP that we are employing. The dimension of the submatrices is 2x2 and the submatrices can be easily extracted:

$$A=\begin{bmatrix}0 & 0\\ \mu & 0\end{bmatrix},\quad B=\begin{bmatrix}-(\lambda+\mu) & \mu\\ 0 & -(\lambda+\mu)\end{bmatrix},\quad C=\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}$$

$$B_1=\begin{bmatrix}-(\lambda+\mu) & \mu\\ \mu & -(\lambda+\mu)\end{bmatrix},\quad B_M=\begin{bmatrix}-\mu & \mu\\ 0 & -\mu\end{bmatrix}$$

Model 4: Server With Erlang-2 Service Times, Breakdown and Repair, Limited Buffer

This model shown in Figure 3.9 combines the characteristics of Model 2 and Model 3. We have a breakdown and repair mechanism just as in Model 2, yet here the production times are distributed with Erlang-2 distribution as is the case in Model 3. To accommodate for these we need three states corresponding to each counting set level. The Markov transition diagram for the process is given in Figure 3.10.

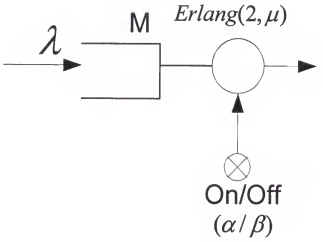

Figure 3.9 Server with Erlang-2 service times, breakdown and repair, limited buffer and its relevant parameters.

This time the state space is the Kroenicker product of the counting set and the set {1,2,B}. This second set is to specify the Erlang stage in the processing when the machine or to specify that the machine is broken. But only two states correspond to the element zero of the counting set. This element has to be handled separately and the QBDP, as we defined it, starts from element 1.

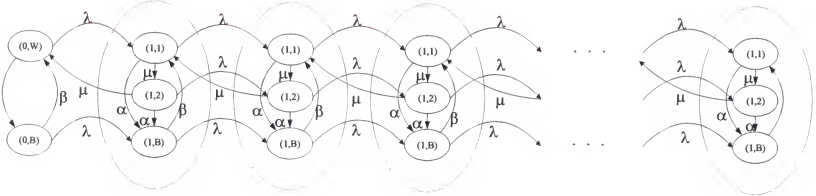

Figure 3.10 Markov transition diagram of Model 4

The steady-state equations for the process are

$$-(\alpha+\lambda)P(0,W)+\beta P(0,B)+\mu P(1,2)=0$$

$$-(\beta+\lambda)P(0,B)+\alpha P(0,W)=0$$

$$-(\alpha+\lambda+\mu)P(1,1)+\lambda P(0,W)+\beta P(1,B)+\mu P(2,2)=0$$

$$-(\alpha+\lambda+\mu)P(1,2)+\mu P(1,1)=0$$

$$-(\beta+\lambda)P(1,B)+\lambda P(0,B)+\alpha P(1,1)+\alpha P(1,2)=0$$

$$-(\alpha+\lambda+\mu)P(i,1)+\lambda P(i-1,1)+\beta P(i,B)+\mu P(i+1,2)=0 \quad for\ 2\le i\le M-1$$

$$-(\alpha+\lambda+\mu)P(i,2)+\lambda P(i-1,2)+\mu P(i,1)=0 \quad for\ 2\le i\le M-1$$

$$-(\beta+\lambda)P(i,B)+\lambda P(i-1,B)+\alpha P(i,1)+\alpha P(i,2)=0 \quad for\ 2\le i\le M-1$$

$$-(\alpha+\mu)P(M,1)+\lambda P(M-1,1)+\beta P(M,B)=0$$

$$-(\alpha+\mu)P(M,2)+\lambda P(M-1,2)+\mu P(M,1)=0$$

$$-\beta P(M,B)+\lambda P(M-1,B)+\alpha P(M,1)+\alpha P(M,2)=0$$

After substituting the values of $P(0,W)$ and $P(0,B)$ obtained from the first two equations into the third and fifth equations, one can obtain the transition rate matrix which is of the form given in Equation (3.2). The submatrices should be defined as

$$A=\begin{bmatrix}0&0&0\\ \mu&0&0\\ 0&0&0\end{bmatrix},\quad B=\begin{bmatrix}-(\alpha+\lambda+\mu)&\mu&\alpha\\ 0&-(\alpha+\lambda+\mu)&\alpha\\ \beta&0&-(\beta+\lambda)\end{bmatrix}$$

$$C=\begin{bmatrix}\lambda&0&0\\ 0&\lambda&0\\ 0&0&\lambda\end{bmatrix},\quad B_1=\begin{bmatrix}-(\alpha+\lambda+\mu)&\mu&\alpha\\ \dfrac{\mu(\beta+\lambda)}{(\alpha+\lambda+\mu)}&-(\alpha+\lambda+\mu)&\dfrac{\alpha\mu}{(\alpha+\lambda+\mu)}+\alpha\\ \beta&0&-(\beta+\lambda)\end{bmatrix}$$

$$B_M = \begin{bmatrix} -(\alpha+\mu) & \mu & \alpha \\ 0 & -(\alpha+\mu) & \alpha \\ \beta & 0 & -\beta \end{bmatrix}$$

.

Model 5: Two Exponential Servers with One Intermediate Limited Buffer

In this model, two exponential servers are present and the buffer is placed between the servers. The typical two-server model in the literature assumes an infinite source for the server, which is stochastically the equivalent of Model 1. The input is a Poisson process. One can assume that there is a second buffer at the input of the system. But the size of this buffer would be limited to one. In following models we will try to illustrate the difficulties faced when a generic sized second buffer is incorporated in the models. The Markov transition diagram for the process is given in Figure 3.12.

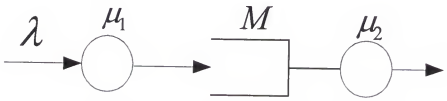

Figure 3.11 Two exponential servers with one intermediate limited buffer and its relevant parameters.

The state space is the Kroenicker product of the counting set and the set {0,1}. The counting set relates to the buffer level as usual. The set {0,1} can be considered as the level of an imaginary first buffer whose size is one; in other words it tells whether there is a part being processed in server one.

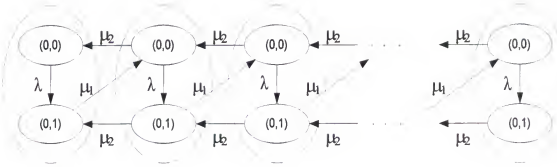

Figure 3.12 Markov transition diagram of Model 5

The steady-state equations for the process are

$$-\lambda P(0,0) + \mu_2 P(1,0) = 0$$

$$-\mu_1 P(0,1) + \lambda P(0,0) + \mu_2 P(1,1) = 0$$

$$-(\lambda + \mu_2) P(i,0) + \mu_1 P(i-1,1) + \mu_2 P(i+1,0) = 0 \qquad \textit{for } 1 \le i \le M-1$$

$$-(\mu_1 + \mu_2) P(i,1) + \lambda P(i,0) + \mu_2 P(i+1,1) = 0 \qquad \textit{for } 1 \le i \le M-1$$

$$-(\lambda + \mu_2) P(M,0) + \mu_1 P(M-1,1) = 0$$

$$-\mu_2 P(M,1) + \lambda P(M,0) = 0$$

This process conforms to the definition of a QBDP that we are employing. Thus all one has to do for formally defining the QBDP is to state the relevant submatrices whose sizes are 2x2:

$$A = \begin{bmatrix} \mu_2 & 0 \\ 0 & \mu_2 \end{bmatrix}, \quad B = \begin{bmatrix} -(\lambda + \mu_2) & \lambda \\ 0 & -(\mu_1 + \mu_2) \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ \mu_1 & 0 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} -\lambda & \lambda \\ 0 & -\mu_1 \end{bmatrix}, \quad B_M = \begin{bmatrix} -(\lambda + \mu_2) & \lambda \\ 0 & -\mu_2 \end{bmatrix}$$

Model 6: Two Exponential Servers With One Intermediate Limited Buffer, Breakdown and Repair

This model is an extension of Model 4. The breakdown and repair mechanisms are added to two basic servers of Model 4. Both servers become non-operational when a breakdown occurs at the server, up to the time that the repair ends. Thus the production process is hindered by this occurrences. These breakdowns and repairs occur independently in both servers. The modeling of this system requires a large number of states for each intermediate buffer level as illustrated in the Markov transition diagram of the process in Figure 3.14.

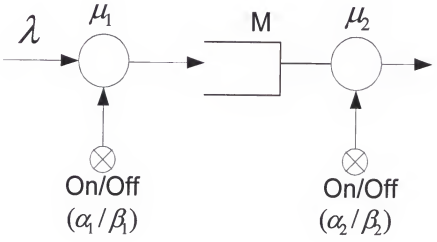

Figure 3.13 Two exponential servers with one intermediate limited buffer, breakdown and repair and its relevant parameters

The state space for this model is the Kroenicker product of four sets: the counting set C–which appears in all the QBDPs–modeling the intermediate buffer level, the set $S_1 = \{W,R\}$ modeling the repair status of the first server, the set $L_1 = \{0,1\}$ modeling

number of parts in server one, and the set $S_2 = \{W,R\}$ modeling the repair status of the second server.

The steady-state equations for the process are

$$-(\lambda+\alpha_1+\alpha_2)P(0,W,0,W)+\beta_1P(0,W,0,R)+\beta_2P(0,R,0,W)+\mu_2P(1,W,0,W)=0$$

$$-(\lambda+\alpha_2+\beta_2)P(0,W,0,R)+\alpha_1P(0,W,0,W)+\beta_2P(0,R,0,R)+\mu_2P(1,W,0,R)=0$$

$$-(\alpha_1+\alpha_2+\mu_1)P(0,W,1,W)+\lambda P(0,W,0,W)+\beta_1P(0,W,1,R)$$
$$+\beta_2P(0,R,1,W)+\mu_2P(1,W,1,W)=0$$

$$-(\alpha_2+\beta_1)P(0,W,1,R)+\lambda P(0,W,0,R)+\alpha_1P(0,W,1,W)$$
$$+\beta_2P(0,R,1,R)+\mu_2P(1,W,1,R)=0$$

$$-(\lambda+\alpha_1+\beta_2)P(0,R,0,W)+\alpha_2P(0,W,0,W)+\beta_1P(0,R,0,R)=0$$

$$-(\lambda+\beta_1+\beta_2)P(0,R,0,R)+\alpha_1P(0,R,0,W)+\alpha_2P(0,W,0,R)=0$$

$$-(\alpha_1+\beta_2+\mu_1)P(0,R,1,W)+\lambda P(0,R,0,W)+\alpha_2P(0,W,1,W)+\beta_1P(0,R,1,R)=0$$

$$-(\beta_1+\beta_2)P(0,R,1,R)+\lambda P(0,R,0,R)+\alpha_1P(0,R,1,W)+\alpha_2P(0,W,1,R)=0$$

$$-(\lambda+\alpha_1+\alpha_2+\mu_2)P(i,W,0,W)+\beta_1P(i,W,0,R)+\beta_2P(i,R,0,W)$$
$$+\mu_2P(i+1,W,0,W)+\mu_1P(i-1,W,1,W)=0 \quad \textit{for } 1\le i\le M-1$$

$$-(\lambda+\alpha_2+\beta_2+\mu_2)P(i,W,0,R)+\alpha_1P(i,W,0,W)+\beta_2P(i,R,0,R)$$
$$+\mu_2P(i+1,W,0,R)=0 \quad \textit{for } 1\le i\le M-1$$

$$-(\alpha_1+\alpha_2+\mu_1+\mu_2)P(i,W,1,W)+\lambda P(i,W,0,W)+\beta_1P(i,W,1,R)$$
$$+\beta_2P(i,R,1,W)+\mu_2P(i+1,W,1,W)=0 \quad \textit{for } 1\le i\le M-1$$

$$-(\alpha_2+\beta_1+\mu_2)P(i,W,1,R)+\lambda P(i,W,0,R)+\alpha_1P(i,W,1,W)$$
$$+\beta_2P(i,R,1,R)+\mu_2P(i+1,W,1,R)=0 \quad \textit{for } 1\le i\le M-1$$

$$-(\lambda+\alpha_1+\beta_2)P(i,R,0,W)+\alpha_2P(i,W,0,W)+\beta_1P(i,R,0,R)$$
$$+\mu_1P(i-1,R,1,W)=0 \quad \textit{for } 1\le i\le M-1$$

$$-(\lambda+\beta_1+\beta_2)P(i,R,0,R)+\alpha_1P(i,R,0,W)+\alpha_2P(i,W,0,R)=0 \quad \textit{for } 1\le i\le M-1$$

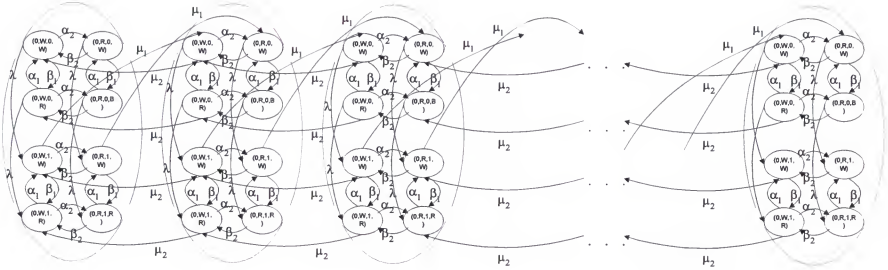

Figure 3.14 Markov transition diagram of Model 6

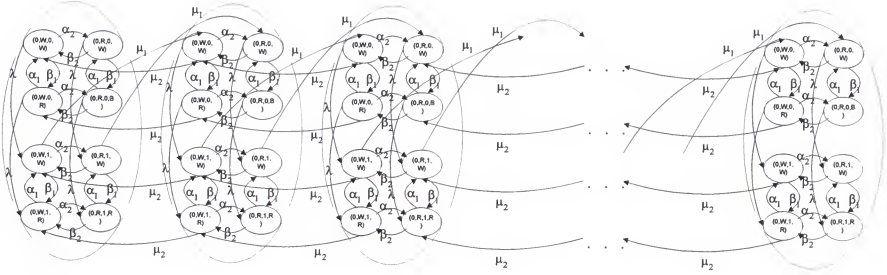

Figure 3.14 Markov transition diagram of Model 6

$$-(\alpha_1+\beta_2+\mu_1)P(i,R,1,W)+\lambda P(i,R,0,W)+\alpha_2 P(i,W,1,W) \\ +\beta_1 P(i,R,1,R)=0 \qquad for\ 1\le i\le M-1$$

$$-(\beta_1+\beta_2)P(i,R,1,R)+\lambda P(i,R,0,R)+\alpha_1 P(i,R,1,W) \\ +\alpha_2 P(i,W,1,R)=0 \qquad for\ 1\le i\le M-1$$

$$-(\lambda+\alpha_1+\alpha_2+\mu_3)P(M,W,0,W)+\beta_1 P(M,W,0,R)+\beta_2 P(M,R,0,W) \\ +\mu_1 P(M-1,W,1,W)=0$$

$$-(\lambda+\alpha_2+\beta_2+\mu_2)P(M,W,0,R)+\alpha_1 P(M,W,0,W)+\beta_2 P(M,R,0,R)=0$$

$$-(\alpha_1+\alpha_2+\mu_2)P(M,W,1,W)+\lambda P(M,W,0,W)+\beta_1 P(M,W,1,R) \\ +\beta_2 P(M,R,1,W)=0$$

$$-(\alpha_2+\beta_1+\mu_2)P(M,W,1,R)+\lambda P(M,W,0,R)+\alpha_1 P(M,W,1,W) \\ +\beta_2 P(M,R,1,R)=0$$

$$-(\lambda+\alpha_1+\beta_2)P(M,R,0,W)+\alpha_2 P(M,W,0,W)+\beta_1 P(M,R,0,R) \\ +\mu_1 P(M-1,R,1,W)=0$$

$$-(\lambda+\beta_1+\beta_2)P(M,R,0,R)+\alpha_1 P(M,R,0,W)+\alpha_2 P(M,W,0,R)=0$$

$$-(\alpha_1+\beta_2)P(M,R,1,W)+\lambda P(M,R,0,W)+\alpha_2 P(M,W,1,W)+\beta_1 P(M,R,1,R)=0$$

$$-(\beta_1+\beta_2)P(M,R,1,R)+\lambda P(M,R,0,R)+\alpha_1 P(M,R,1,W)+\alpha_2 P(M,W,1,R)=0$$

This process conforms to the definition of a QBDP that we are employing. Thus all one has to do for formally defining the QBDP is to state the relevant submatrices whose sizes are 8x8:

$$A=\begin{bmatrix}\mu_2&0&0&0&0&0&0&0\\0&\mu_2&0&0&0&0&0&0\\0&0&\mu_2&0&0&0&0&0\\0&0&0&\mu_2&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\end{bmatrix},\quad C=\begin{bmatrix}0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\\mu_1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&\mu_1&0&0&0\\0&0&0&0&0&0&0&0\end{bmatrix}$$

$$B=\begin{bmatrix}-(\lambda+\alpha_1+\alpha_2+\mu_1)&\alpha_1&\lambda&0&\alpha_2&0&0&0\\\beta_1&-(\lambda+\alpha_1+\beta_1+\mu_1)&0&\lambda&0&\alpha_2&0&0\\0&0&-(\alpha_1+\alpha_2+\mu_1+\mu_2)&\alpha_1&0&0&\alpha_2&0\\0&0&\beta_1&-(\alpha_2+\beta_1+\mu_1)&0&0&0&\alpha_2\\\beta_2&0&0&0&-(\lambda+\alpha_1+\beta_2)&\alpha_1&\lambda&0\\0&\beta_2&0&0&\beta_1&-(\lambda+\beta_1+\beta_2)&0&\lambda\\0&0&\beta_2&0&0&0&-(\alpha_1+\beta_2+\mu_1)&\alpha_1\\0&0&0&\beta_2&0&0&\beta_1&-(\beta_1+\beta_2)\end{bmatrix}$$

$$B_0=\begin{bmatrix}-(\lambda+\alpha_1+\alpha_2)&\alpha_1&\lambda&0&\alpha_2&0&0&0\\\beta_1&-(\lambda+\alpha_1+\beta_1)&0&\lambda&0&\alpha_2&0&0\\0&0&-(\alpha_1+\alpha_2+\mu_1)&\alpha_1&0&0&\alpha_2&0\\0&0&\beta_1&-(\alpha_2+\beta_1)&0&0&0&\alpha_2\\\beta_2&0&0&0&-(\lambda+\alpha_1+\beta_2)&\alpha_1&\lambda&0\\0&\beta_2&0&0&\beta_1&-(\lambda+\beta_1+\beta_2)&0&\lambda\\0&0&\beta_2&0&0&0&-(\alpha_1+\beta_2+\mu_1)&\alpha_1\\0&0&0&\beta_2&0&0&\beta_1&-(\beta_1+\beta_2)\end{bmatrix}$$

$$B_M=\begin{bmatrix}-(\lambda+\alpha_1+\alpha_2+\mu_1)&\alpha_1&\lambda&0&\alpha_2&0&0&0\\\beta_1&-(\lambda+\alpha_1+\beta_1+\mu_1)&0&\lambda&0&\alpha_2&0&0\\0&0&-(\alpha_1+\alpha_2+\mu_1)&\alpha_1&0&0&\alpha_2&0\\0&0&\beta_1&-(\alpha_2+\beta_1+\mu_1)&0&0&0&\alpha_2\\\beta_2&0&0&0&-(\lambda+\alpha_1+\beta_2)&\alpha_1&\lambda&0\\0&\beta_2&0&0&\beta_1&-(\lambda+\beta_1+\beta_2)&0&\lambda\\0&0&\beta_2&0&0&0&-(\alpha_1+\beta_2)&\alpha_1\\0&0&0&\beta_2&0&0&\beta_1&-(\beta_1+\beta_2)\end{bmatrix}$$

Model 7: Server With Erlang-N Service Times, Limited Buffer

In Model 3, we have studied the case when the server has Erlang-2 service times. In this model we extend the model to Erlang-N service times. The number of states in this model increases both with M and N. Yet our solution procedure does not depend on the size of the counting space which is the buffer size here. But the larger the Erlang parameter is, the larger the submatrices are. Thereby the computational effort increases with the parameter N. The Markov transition diagram of the process is given below:

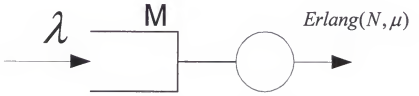

Figure 3.15 Server with Erlang-N service times, limited buffer and its relevant parameters

The state space for this model is the Kroenicker product of two sets. The counting set C that corresponds to the buffer level and the Erlang stage set S. The Erlang stage set itself, is also a counting set. Thus one can two different matrix representations for the QBDP: one using the buffer level as the counting set of the QBDP and one using the Erlang stage set. But since there is a transition from the Erlang stage N to Erlang stage 1, the choice of Erlang stage set would not yield a QBDP according to our definition.

The steady-state equations for the process are

$$-\lambda P(0) + \mu P(1, N) = 0$$

$$-(\lambda+\mu)P(1,1)+\lambda P(0)+\mu P(2,N)=0$$

$$-(\lambda+\mu)P(1,j)+\mu P(1,j-1)=0 \qquad for\ \ 1\le j\le N$$

$$-(\lambda+\mu)P(i,1)+\lambda P(i-1,1)+\mu P(i,N)=0 \qquad for\ \ 2\le i\le M-1$$

$$-(\lambda+\mu)P(i,j)+\lambda P(i-1,j)+\mu P(i,j-1)=0 \quad for\ \ 2\le i\le M-1\ \ and\ \ 2\le j\le N$$

$$-\mu P(M,1)+\lambda P(M-1,1)=0$$

$$-\mu P(M,j)+\lambda P(M-1,j)+\mu P(M,j-1)=0 \qquad 2\le j\le N$$

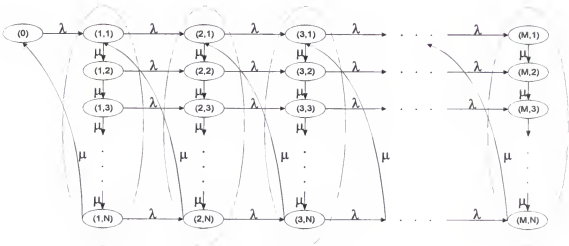

Figure 3.16 Markov transition diagram of Model 7

After substituting the value of $P(0)$ obtained from the equation one can obtain the transition rate matrix for this process. The transition rate matrix conforms to the definition of the QBDPs as one could expect. This time the submatrices employed are of size NxN:

$$A = \begin{bmatrix} 0 & 0 & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & 0 \\ \mu & 0 & . & . & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -(\lambda+\mu) & \mu & & & \\ & -(\lambda+\mu) & . & & \\ & & . & . & \\ & & & . & \mu \\ & & & & -(\lambda+\mu) \end{bmatrix}$$

$$C = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & . & & \\ & & & . & \\ & & & & \lambda \end{bmatrix}, \quad B_1 = \begin{bmatrix} -(\lambda+\mu) & \mu & & & \\ & -(\lambda+\mu) & . & & \\ & & . & . & \\ & & & . & \mu \\ \mu & & & & -(\lambda+\mu) \end{bmatrix}$$

$$B_M = \begin{bmatrix} -\mu & \mu & & & \\ & -\mu & . & & \\ & & . & . & \\ & & & . & \mu \\ & & & & -\mu \end{bmatrix}$$

If one wanted to use Erlang stage set as the counting set, he would not get the a QBDP in the strict sense of the word, but something quite close to it. And our matrix polynomial solution procedure would still be applicable since the deviation from the QBDP definition would only occur in the boundary equations. This is due to the fact that the matrix polynomial solution procedure is based on the structure of the inner equations. It does not preclude a deviation from the definition in boundary equations.

The transition rate matrix obtained with such a choice would be of the form

$$R = \begin{bmatrix} B_0 & C & & & & & X_M \\ A & B & C & & & & \\ & A & B & . & & & \\ & & A & . & . & & \\ & & & & . & . & \\ & & & & . & B & C \\ & & & & A & B & C \\ X_0 & & & & & A & B_M \end{bmatrix} \tag{3.3}$$

This one is a generalization of the transition rate matrix given in equation (3.2). We see that two impurities are allowed at the corners of the matrix.

Once this transition rate matrix is defined, one can write the transition rate matrix of the process with the Erlang stage used as the counting set within this platform. Indeed to define the submatrices in the (3.3) would suffice:

$$B = \begin{bmatrix} -(\lambda+\mu) & \lambda & & & & \\ & -(\lambda+\mu) & . & . & & \\ & & & . & . & \\ & & & & \lambda & \\ & & & & -(\lambda+\mu) & \lambda \\ & & & & & -\mu \end{bmatrix}, \quad B_1 = B$$

$$B_M = \begin{bmatrix} -\mu & \mu & & & & \\ & -\mu & . & . & & \\ & & & . & . & \\ & & & & \mu & \\ & & & & -\mu & \mu \\ & & & & & -\mu \end{bmatrix}, \quad C = \begin{bmatrix} \mu & & & \\ & \mu & & \\ & & . & \\ & & & . \\ & & & \mu \end{bmatrix}, \quad A = \underline{0}$$

$$X_1 = \begin{bmatrix} \mu & & & & & \\ 0 & 0 & & & & \\ & \mu & 0 & & & \\ & & \mu & . & & . \\ & & & . & . & \\ & & & & 0 & \\ & & & & \mu & 0 \end{bmatrix}, \quad X_M = \underline{0}$$

Model 8: Two Exponential Servers with Two Limited Buffers

This model is an extension on the Model 5. The first server did not have a buffer for the arrivals. Thereby those arrivals that occurred while the server was occupied were lost. In this model, those arrivals are stored in a buffer whose size is M_1. This model is quite important because it is a tandem queue, which happens to form the basic building block of many models for modern production systems. Yet, a theoretically sound methodology has not developed to tackle the stochastic analysis of this queue class up to this point.

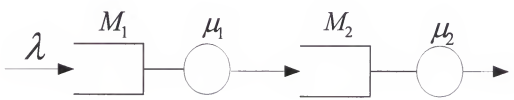

Figure 3.17 Two exponential servers with two limited buffers and their relevant parameters

One way of doing it is to use a larger state-space to incorporate the second queue. Though this one is a valid option, since the computational effort increases with the size of

the state space, it is not viable for many real models. Moreover it is not conducive to general results from which one can obtain an idea about the behavior of such systems.

Although we have attempted many times, to tackle this problem ourselves, we have not been able to bring about a solution. Within the platform of matrix polynomial solution procedure, one can solve QBDPs with an effort independent of the size of the counting set. Yet, for the tandem queues one does not have only one counting set but many of them. For this model with two tandem queues we have to counting sets, each corresponding to the buffer levels of the two servers.

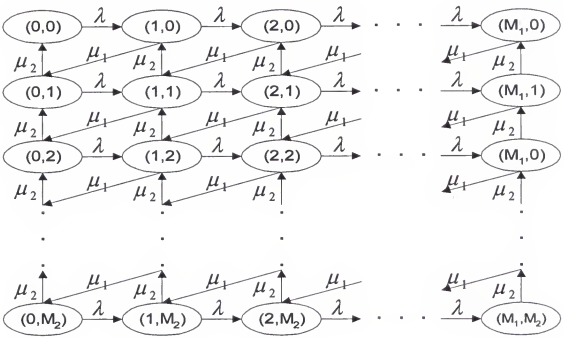

Figure 3.18 Markov transition diagram of Model 8

Since the methodology we employ only allows us to exploit the structure for a single counting set, we cannot make use of the second counting set. One point to

consider is the choice of the principal counting set of the QBDP. The logical strategy for this is to choose the one with higher cardinality. Then the computational effort that relates to the solution is decreased.

To illustrate better the nature of the model, let us present the Markov transition diagram pertaining to the model in Figure 3.18. One can clearly observe by examining the diagram that the state space is the Kroenicker product if two counting sets: the first buffer level set $C_1 = \{1,2,3, \ldots, M_1\}$ and the second buffer level set $C_2 = \{1,2,3, \ldots, M_2\}$. The first index of the states is element of C_1 and the second index of C_2. We also observe that there are $M_1.M_2$ states in the model.

The steady-state equations for the process are

$$-\lambda P(0,0) + \mu_2 P(0,1) = 0$$

$$-(\lambda + \mu_1)P(i,0) + \lambda P(i-1,0) + \mu_2 P(i,1) = 0 \qquad for\ 1 \le i \le M_1 - 1$$

$$-\mu_1 P(M_1,0) + \lambda P(M_1 - 1,0) + \mu_2 P(M_1,1) = 0$$

$$-(\lambda + \mu_2)P(0,j) + \mu_1 P(1,j-1) + \mu_2 P(0,j+1) = 0 \qquad for\ 1 \le j \le M_2 - 1$$

$$-(\lambda + \mu_2)P(0,M_2) + \mu_1 P(1,M_2 - 1) = 0$$

$$-(\lambda + \mu_1 + \mu_2)P(i,j) + \lambda P(i-1,j) + \mu_1 P(i+1,j-1) + \mu_2 P(i,j+1) = 0$$

$$for\ 1 \le i \le M_1 - 1\ \ 1 \le j \le M_2 - 1$$

$$-(\lambda + \mu_2)P(i,M_2) + \lambda P(i-1,M_2) + \mu_1 P(i+1,M_2 - 1) = 0 \qquad for\ 1 \le j \le M_2 - 1$$

$$-(\mu_1 + \mu_2)P(M_1,j) + \lambda P(M_1 - 1,j) + \mu_2 P(M_1,j+1) = 0 \qquad for\ 1 \le j \le M_2 - 1$$

$$-\mu_2 P(M_1,M_2) + \lambda P(M_1 - 1,M_2) = 0$$

Lucidly, this process conforms to the definition of the QBDP. Moreover, one has two choices for the counting set. If the set C_1 is elected as the counting set, the

submatrices that define the QBDP are of the dimension $M_2 x M_2$. These relevant submatrices are

$$A = \begin{bmatrix} 0 & \mu_1 & & & & \\ & 0 & \mu_1 & & & \\ & & 0 & . & & \\ & & & . & . & \\ & & & & . & \mu_1 \\ & & & & & 0 \end{bmatrix}, \quad C = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & . & & \\ & & & . & \\ & & & & \lambda \end{bmatrix}$$

$$B = \begin{bmatrix} -(\lambda+\mu_1) & & & & & & & \\ \mu_2 & -(\lambda+\mu_1+\mu_2) & & & & & & \\ & \mu_2 & -(\lambda+\mu_1+\mu_2) & & & & & \\ & & & \mu_2 & . & & & \\ & & & & . & . & & \\ & & & & & . & -(\lambda+\mu_1+\mu_2) & \\ & & & & & & \mu_2 & -(\lambda+\mu_2) \end{bmatrix}$$

$$B_0 = \begin{bmatrix} -\lambda & & & & & & \\ \mu_2 & -(\lambda+\mu_2) & & & & & \\ & \mu_2 & -(\lambda+\mu_2) & & & & \\ & & \mu_2 & . & & & \\ & & & . & . & & \\ & & & & . & -(\lambda+\mu_2) & \\ & & & & & \mu_2 & -(\lambda+\mu_2) \end{bmatrix}$$

$$B_{M_1} = \begin{bmatrix} -\mu_1 & & & & & & \\ \mu_2 & -(\mu_1+\mu_2) & & & & & \\ & \mu_2 & -(\mu_1+\mu_2) & & & & \\ & & \mu_2 & . & & & \\ & & & . & . & & \\ & & & & . & -(\mu_1+\mu_2) & \\ & & & & & \mu_2 & -(\mu_2) \end{bmatrix}$$

If the set C_2 is elected as the counting set, the submatrices that define the QBDP are of the dimension $M_1 x M_1$. These relevant submatrices are

$$A = \begin{bmatrix} \mu_2 & & & & \\ & \mu_2 & & & \\ & & . & & \\ & & & . & \\ & & & & \mu_2 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & & & & & \\ \mu_2 & 0 & & & & \\ & \mu_2 & . & & & \\ & & . & . & & \\ & & & . & 0 & \\ & & & & \mu_2 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} -(\lambda+\mu_2) & \lambda & & & & & & \\ & -(\lambda+\mu_1+\mu_2) & \lambda & & & & & \\ & & -(\lambda+\mu_1+\mu_2) & . & & & & \\ & & & . & . & & & \\ & & & & . & \lambda & & \\ & & & & & -(\lambda+\mu_1+\mu_2) & \lambda & \\ & & & & & & -(\mu_1+\mu_2) \end{bmatrix}$$

$$B_0 = \begin{bmatrix} -\lambda & \lambda & & & & & \\ & -(\lambda+\mu_1) & \lambda & & & & \\ & & -(\lambda+\mu_1) & . & & & \\ & & & . & . & & \\ & & & & . & \lambda & \\ & & & & & -(\lambda+\mu_1) & \lambda \\ & & & & & & -\mu_1 \end{bmatrix}$$

$$B_{M_2} = \begin{bmatrix} -(\lambda+\mu_2) & \lambda & & & & & \\ & -(\lambda+\mu_2) & \lambda & & & & \\ & & -(\lambda+\mu_2) & . & & & \\ & & & . & . & & \\ & & & & . & \lambda & \\ & & & & & -(\lambda+\mu_2) & \lambda \\ & & & & & & -\mu_2 \end{bmatrix}$$

The interesting phenomena that one should observe is that the block tridiagonal structure that is the insignia of the QBDP is also exhibited in each of the submatrices that

we have presented. Indeed, it is possible to show that this phenomenon is projected further for tandem queues having more than servers.

One can relate this phenomenon to fractals. The same kind of macro-structure appears in the elements of the whole. As one gets to examine these submatrices the same structure exhibits itself. And the amount that this self-similarness occurs is the number of stations–and thereby the number of buffers–in the tandem queue.

Although we were unable to exploit this marvelous structure in order to obtain the nullspace of the transition rate matrix, it can be quite useful in generating the transition rate matrix for a given process in an efficient fashion.

Thus, one can conclude that there is a reasonable amount of material in this area to justify future research. Although we have not been able to exploit the structure we have described, it is certainly still an effort worthy of the attention of future researchers. Thus, here we relegate this effort to them.

CHAPTER 4
A GENERAL THEORY ON SPECTRAL PROPERTIES OF STATE-HOMOGENEOUS FINITE-STATE QUASI-BIRTH-DEATH PROCESSES

4.1 Introduction

In this chapter, we present a general theory for the spectral properties of state-homogeneous finite-state Quasi-Birth-Death Processes (QPDP). This one is the main theoretical contribution of the thesis. Our purpose is to depict the general structure of the state-homogeneous finite-state QBDPs, which are quite pervasive in the domain of stochastic modeling. Although much previous research has been done on the subject of QBDPs, matrix polynomial approaches have only been recently applied. Consequently, the spectral properties of these processes, which are of fundamental importance in the application of matrix polynomial methods, have not been thoroughly investigated up to this point. This chapter develops the theory which is to constitute the general foundation for the application of matrix polynomial methods.

It is known (Yeralan and Muth 1987) that a state-homogenous finite-state quasi-birth-death process gives rise to a transition rate matrix having a nested block tridiagonal structure.

$$R = \begin{bmatrix} B_0 & C & & & & & \\ A & B & C & & & & \\ & A & B & C & & & \\ & & . & . & . & & \\ & & & . & . & . & \\ & & & & A & B & C \\ & & & & & A & B_M \end{bmatrix} \tag{4.1}$$

The matrix R, the transition rate matrix, carries all the information needed to characterize a Markov process. Using this matrix, one can evaluate different quantities of interest for a given process. One point of interest is its evolution in time referred to as the transient behavior. The transition rate matrix along with an initial condition vector is sufficient to determine the transient behavior. Yet, although one may have all the information needed in hand, the actual evaluation of the transient behavior may be computationally intractable, or just too cumbersome due to the large size of the matrix R.

For most analysis, one is interested in the steady-state behavior of the process. This means one would like to learn where the process would converge after it settles. Since the transient behavior is being overlooked, this kind of analysis is more easily done. All one has to do to obtain the steady-state probabilities of an irreducible Markov process is the find the null-space of the matrix R and then choose the only element of the null-space having its components adding to one. For any irreducible Markov process, there is always a unique element satisfying this property.

In the case of state-homogeneous finite-state QBDPs, which belong to the general category of Markov processes, some additional structure that is prone to further exploitation is present. Most visibly, the rate matrix has a nested tridiagonal structure. Thus a method needs to be devised in order to take advantage of this structure. And here

we argue that the use of matrix polynomial technique is quite propitious. That was first proposed by Tan (1994).

4.2 The Matrix Polynomial Solution Procedure

All quantities of interest for the evaluation of steady-state probabilities of Markov process is within the transition rate matrix, R. Moreover, the matrix R for the QBDPs has the block tridiagonal form as depicted in the equation (4.1). The steady-state probability vector, π–towards which our effort is geared, –is known to satisfy the following equations.

$$\pi R = \underline{0} \tag{4.2}$$

$$\pi \underline{1}^T = 1 \tag{4.3}$$

where R is the transition rate matrix, $\underline{1}$ = (1,1,…,1), and $\underline{0}$ = (0,0,…,0). As we have already stated this probability vector is unique provided that R is irreducible. Any vector g satisfying the equation (4.2) is called a nonnormalized steady-state probability vector. Let $(g_0, g_1, \ldots, g_M)$ be a partitioning of the vector g in such a way that the size of each partition matches the dimension of the corresponding partition in R as manifested in equation (4.1). Thus we can state

$$gR = \underline{0} \tag{4.4}$$

Then using the partition notation we can rewrite equation (4.4).

$$g_0 B_0 + g_1 A = \underline{0} \tag{4.5}$$

$$g_{i-1} C + g_i B + g_{i+1} A = \underline{0} \quad for\ i = 1,2,\ldots, M-1 \tag{4.6}$$

$$g_{M-1}C + g_M B_M = \underline{0} \tag{4.7}$$

Equations (4.5) and (4.7) are named as the boundary equations and equation (4.6) as the interior equation. The interior equation is actually repeated M-1 times and can actually be classified as a difference equation. We propose to exploit this structure of the interior equation applying the matrix polynomial solution procedure.

It has been shown (Lancaster, Gohlberg and Rodman, 1982) that this matrix difference equation has solutions in the following form:

$$g_i = \lambda^i e \quad for\ i = 0,1,2,\ldots, M \tag{4.8}$$

where λ is scalar and e is a vector of the same dimension as g_i. In order to find all solutions of this family we substitute the proposed solution into the equation (4.6) and thereby obtain

$$\lambda^{i-1} eC + \lambda^i eB + \lambda^{i+1} eA = \underline{0} \tag{4.9}$$

For the values of λ different then zero this equation simplifies to

$$\lambda^2 eA + \lambda eB + eC = \underline{0} \tag{4.10}$$

Now, we define the matrix polynomial $L(\lambda)$ in order to formalize our problem within the framework of matrix polynomial theory:

$$L(\lambda) = \lambda^2 A + \lambda B + C \tag{4.11}$$

Consequently, our goal–finding all vectors g satisfying equation (4.6)–can be reformulated as finding the eigenvalue-eigenvector pairs that belongs to the matrix polynomial $L(\lambda)$. We can also express the equation (4.10) using the new notation as

$$eL(\lambda) = \underline{0} \tag{4.12}$$

The set of all eigenvalue-eigenvector pairs (λ_i, e_i) of $L(\lambda)$ yields us a set of linearly independent solution vectors which spans the general solution space of the equation (4.6). The eigenvalues are the roots of the characteristic polynomial $\det(L(\lambda)) = 0$. One can readily observe this characteristic equation is of degree 2n and thereby has 2n solutions where n is the size of one dimension for each partition of R. Since each eigenvalue is an element of the extended complex numbers set–the extension is for the introduction of ∞ to the complex numbers set, –we can propose such an ordering of the eigenvalues:

$\lambda_1, \ldots, \lambda_Z = 0$

$\lambda_{Z+1}, \ldots, \lambda_{Z+N}$ finite and nonzero

$\lambda_{Z+N+1}, \ldots, \lambda_{2n} = \infty$

where Z is the number of zero eigenvalues and N is the number of finite and nonzero eigenvalues. Muller's method is appropriate to obtain the characteristic equation, which is of at most degree 2n, and subsequently to find each eigenvalue. One should be warned that in this scheme, in the presence of multiplicity in eigenvalues, each instance of a multiple eigenvalue is counted as a separate one. Now we can observe that Z is the dimension of the null space of C; and 2n-Z-N is the dimension of the null space of A. By the same token, the eigenvectors belonging to the null eigenvalues are the vectors spanning the null space of C; and those belonging to the infinite eigenvalues are the vectors spanning the null space of A.

When we express the general solution using the eigenvalue-eigenvector pairs we get

$$g_i = \sum_{j=Z+1}^{Z+N} w_j \lambda_j^i e_j \quad for\ i = 1,2,\ldots, M-1 \tag{4.13}$$

The equations for g_0 and g_M have to be expressed separately. The reason for this is that the components of the solution corresponding to the eigenvalues at 0 contribute only to g_0; and similarly the components corresponding to the eigenvalues at the infinity contribute only to g_M:

$$g_0 = \sum_{j=1}^{N+Z} w_j e_j \tag{4.14}$$

$$g_M = \sum_{j=N}^{N+Z} w_j \lambda_j^M e_j + \sum_{j=N+Z+1}^{2n} w_j e_j \tag{4.15}$$

In all these equations w_j is the weight associated with component of the solution corresponding to the j-th eigenvalue-eigenvector pair. By changing the weights we can obtain any particular solution.

But all these expressions actually are only correct under the assumption that for each eigenvalue with multiplicity higher than one, there are as many linearly independent genuine eigenvectors belonging to that eigenvalue as the multiplicity of the eigenvector. When this is not the case, the generalized eigenvectors need to be introduced along with Jordan canonical forms.

In order to express the solution within the framework of Jordan pairs, let l be the number of distinct finite eigenvalues of $L(\lambda)$ and consider an ordering of the distinct eigenvalues from 1 to l. Then we can define the a finite Jordan pair (X_F, J_F) of $L(\lambda)$ as

$X_F = [X(\lambda_1), X(\lambda_2), X(\lambda_3), \ldots, X(\lambda_l)]$

$J_F = \text{diag}[J(\lambda_1), J(\lambda_2), J(\lambda_3), \ldots, J(\lambda_l)]$

where $(X(\lambda_j), J(\lambda_j))$ is a Jordan pair for every finite eigenvalue, λ_j, of $L(\lambda)$. In order to express the general solution of equation (4.12), we also need the Jordan pair that belongs to the eigenvalue at the infinity. (X_∞, J_∞) is how this pair is noted and this pair is called the infinite Jordan pair of $L(\lambda)$. One should notice that all the elements of J_∞ are either zero or one, the diagonal elements that would be the eigenvalues in a typical matrix in Jordan form are all one. Then the general solution in its correct form can be expressed as

$$g_i = WJ_F^i X_F \quad for\, i = 0,1,2,\ldots, M-1 \tag{4.16}$$

$$g_M = WJ_F^M X_F + W_\infty J_\infty X_\infty \tag{4.17}$$

where W is a row vector with Z+N elements having as component the weights corresponding to the part of the solution due to finite eigenvalues; and W_∞ is a row vector with 2n-Z-N elements having as component the weights corresponding to the part of the solution due to the eigenvalue at infinity.

Now with the equations (4.16) and (4.17) we have the correct version of the general solution to the matrix difference equation, which is equation (4.5). That means any solution can be expressed as a special case of this one just by adjusting the weights.

Yet, this closed form is not always useful for our purposes. Thus, we will soon present the general solution in a more explicit form, which is instrumental in showing the properties of the QBDPs. Some new notation is needed for this endeavor.

Let ρ be the number of the finite genuine eigenvalues or in other words the number of Jordan canonical blocks in the matrix J_F. Let ρ_∞ be the number of genuine eigenvalues at infinity that is the number of Jordan canonical blocks in the matrix J_∞. Let the series $(\lambda_j / j=1,2,\ldots,\rho+\rho_\infty)$ be a non-decreasing sequence of those eigenvalues with the eigenvectors at infinity at the end of the sequence.

$$g_i = \sum_{j=1}^{\rho} w_j \left[J(\lambda_j)\right]^i X(\lambda_j) + \delta(M,i) \sum_{j=\rho}^{\rho+\rho_\infty} w_j J_\infty(\lambda_j) X_\infty(\lambda_j)$$

$$\text{for i=0,1,2,\ldots,M} \tag{4.18}$$

where w_j is a row vector of the size of Jordan block belonging to λ_j and $\delta(i,j)$ be the indicator function.

It is possible to further expand this expression, since we know the closed-form expression for the powers of the Jordan canonical forms:

$$g_i = \sum_{j=1}^{\rho} \sum_{k=1}^{s(\lambda_j)} w_{(j,k)} \sum_{l=\max(i-k+1,0)}^{i} \binom{i}{l} \lambda_j^l e_{(j,k-i+l)} + \delta(M,i) \sum_{j=\rho}^{\rho+\rho_\infty} \sum_{k=1}^{s(\lambda_j)} w_{(j,k)} \sum_{l=\max(i-k+1,0)}^{i} \binom{i}{l} e_{(j,k-i+l)}$$

$$\text{for i=1,2,\ldots,M} \tag{4.19}$$

where $s(\lambda_j)$ is the length of the generalized eigenvector cycle, or in other words, the size of the Jordan canonical form corresponding to the eigenvector, λ_j; $e_{(j,k)}$ is the k^{th} generalized eigenvector of the generalized eigenvector cycle corresponding to the eigenvector, λ_j; and $w_{(j,k)}$ is the weight that corresponds to the k^{th} generalized eigenvector of the generalized eigenvector cycle corresponding to the eigenvector, λ_j. One should observe in the equation (4.19) that the terms that belong to the free variable $w_{(j,k)}$ not only include the generalized eigenvector corresponding to this term, but also generalized eigenvectors that are preceding it in the same generalized eigenvector cycle. This actually foreshadows some properties that are going be presented in the next section.

The general solution presented has 2n variables that can be set freely. Using the physics jargon, it has 2n degrees of freedom. This freedom is needed since the particular

solution that is of interest needs to satisfy the equations (4.5) and (4.7) which constitute a set of linearly independent equation system with rank 2n-1, along with the normalizing condition. The fact that the rank is 2n-1 is quite critical since if the rank were 2n, the only possible solution of the system, which is homogeneous, would be the trivial solution, being the zero vector. Moreover, since the zero vector solution would always be in contradiction with the normalizing condition, the system would be without a solution. Yet, just as we would expect, the system of equation has a linear dependence. This dependence is also going to be shown in conjunction with the spectral properties of the process in the next section.

4.3 Spectral Properties of State-Homogeneous Finite-State Quasi-Birth-Death Processes

In this section we are going to delve in the spectral properties state-homogeneous finite-state QBDPs manifest. This section will directly be founded on the solution procedure developed in Section 4.2. We begin this treatment of the subject by presenting two readily justifiable assumptions.

Assumption 4.1: The process is ergodic. That is the transition rate matrix R is irreducible.

Assumption 4.2: The matrix (A+B+C) is an irreducible transition rate matrix.

Assumption 4.1 is a standard assumption in the analysis of Markov processes, since if this assumption did not hold, one could always decompose the process into decoupled subprocesses, and apply the same analysis on these subprocesses.

Assumption 4.2 makes certain that interior equations of the process do not yield to any kind of decoupling. The fact that (A+B+C) is a transition rate matrix actually needs to be proven. This can be done by making use of Property 4.2 that will be presented later and the fact that the only negative entries are the diagonal entries of the matrix.

Let the state space of the process be the Kroenicker product the count set $C=\{e\in N/\ 0\le e\le M\}$ and the internal-state-space whose cardinality is the size of all the aforementioned submatrices. It is known that only the diagonal element of B are negative and all the other element of matrices A, B, and C are nonnegative. Since all coupling between the internal states manifest themselves as positive entries in the submatrices A, B and C these couplings would all be preserved in the stochastic matrix, A+B+C. Thus if the mentioned matrix is not an irreducible transition rate matrix, there would be some kind of decoupling in the internal-state-space. That means from a certain group of elements of the space, there would be no transition to another group.

It is also possible to show this decoupling on the elements of the solution procedure. If (A+B+C) is not an irreducible transition rate matrix then the matrix polynomial $L(\lambda)$ would have eigenvalue and eigenvector pairs can be partitioned among the subprocesses. Thus, one could actually work separately with these subprocesses.

One should also make a note that the Assumption 4.1 does not imply Assumption 4.2. Even when A+B+C corresponds to a reducible Markov chain, the full chain may be irreducible due to a coupling at the boundary states–the states corresponding to 0 an M in

the counting set, C. When this one is the case, one can work with two separate processes up to the application of boundary processes, and consequently reduce the computational effort that needs to be exerted.

Property 4.1:

$A\underline{1}^T \neq \underline{0}$ and $C\underline{1}^T \neq \underline{0}$

Since both A and C consist of nonnegative elements the sum of row elements can be zero only if all the elements of corresponding rows are zero. This means for the Property 4.1 not to hold, either A or C would have to be a zero matrix which would in turn render R a reducible transition rate matrix. Since this would contradict Assumption 4.1, Property 4.1 always holds.

Property 4.2:

$$A\underline{1}^T + B\underline{1}^T + C\underline{1}^T = \underline{0}^T \tag{4.20}$$

This property directly follows the fact that R is a transition rate matrix. It is known that the elements in each row of R has to add to zero, or $R\underline{1}^T = \underline{0}^T$. When the R is written in terms of the submatrices that belong to it, one obtains

$$\begin{bmatrix} B_0\underline{1}^T + C\underline{1}^T \\ A\underline{1}^T + B\underline{1}^T + C\underline{1}^T \\ \vdots \\ A\underline{1}^T + B\underline{1}^T + C\underline{1}^T \\ A\underline{1}^T + B_M\underline{1}^T \end{bmatrix} = \begin{bmatrix} \underline{0}^T \\ \underline{0}^T \\ \vdots \\ \underline{0}^T \\ \underline{0}^T \end{bmatrix} \tag{4.21}$$

The interior rows all yield Property 4.2. Moreover, two more relations are obtained from the boundary rows:

$$B_0 \underline{1}^T + C\underline{1}^T = \underline{0}^T \tag{4.22}$$

$$A\underline{1}^T + B_M \underline{1}^T = \underline{0}^T \tag{4.23}$$

Property 4.3:

One is always an eigenvalue of $L(\lambda)$.

By Property 4.2 $(A+B+C)\underline{1}^T = \underline{0}^T$. This can also be written as $L(1)\underline{1}^T = \underline{0}^T$. This means one is an eigenvalue of $L(\lambda)$ and $\underline{1}$ is a right-eigenvector corresponding to it.

Property 4.4:

$$g_i C\underline{1}^T = g_{i+1} A\underline{1}^T \qquad \text{for } i = 0, 1,2, \ldots, M\text{-}1 \tag{4.24}$$

This equation is called the balance of flow equations for the QBDPs. It is a generalization of the balance of flow equations for normal birth-death processes. This equation simply states the fact that in the equilibrium, the rate at which the probability of being in the states corresponding to the i^{th} element of the counting set is transformed to the probability of being in the states corresponding to the $(i+1)^{th}$ element of the counting set, is equal to the rate the reverse occurs. This transformation that conserves the quantities can be likened to a flow, when the liquid metaphor is used for the probabilities.

This property is proven by mathematical induction. By multiplying both sides of the equation (4.5) with the additive vector, $\underline{1}^T$, from the right, one obtains $g_0 B_0 \underline{1}^T + g_1 A\underline{1}^T = \underline{0}^T$. Using (4.22) along this one, one reaches

$$g_0 C\underline{1}^T = g_1 A\underline{1}^T \tag{4.25}$$

Thereby the property holds for i=0. Now one should assume that the property holds for i=k in order to use induction argument, i.e.,

$$g_k C\underline{1}^T = g_{k+1} A\underline{1}^T \tag{4.26}$$

When equation (4.20) is used to substitute $C\underline{1}$ in the equation (4.6) for i=k+1, the following is obtained

$$g_k C\underline{1}^T - g_{k+1}(A\underline{1}^T + C\underline{1}^T) + g_{k+2} A\underline{1}^T = \underline{0}$$

Substituting (4.26) in this expression yields

$$g_{k+1} C\underline{1}^T = g_{k+2} A\underline{1}^T$$

Thereby the identity holds for i=k+1 given that it holds for i=k. Now by using the induction argument we prove that Property 4.4 is always true.

Theorem 4.1:

$$\lambda_j e_{(j,1)} A\underline{1}^T = e_{(j,1)} C\underline{1}^T \tag{4.27}$$

for all eigenvalues λ_j different from one.

The equation (4.27) is referred as a Balance Equation in Component Form (BECF) by Tan who first proposed and proved it. Theorem 4.1 states that BECF hold for all eigenvalue-eigenvector pairs satisfying the matrix polynomial equation (4.12) except for those with eigenvalue one. The case of those with eigenvalue one will be investigated separately. Also, one should notice that Theorem 4.1 does not state anything about the generalized eigenvectors that may be present in the solution due to multiplicity in the roots of characteristic equation.

Proof:

Let $(\lambda_j, e_{(j,1)})$ be an eigenvalue-eigenvector pair as defined for the equation (4.19). This eigenvalue, λ_j, may or may not correspond to a generalized eigenvector cycle. In either case the first element of the cycle, the genuine eigenvector exists. Since this pair is a solution of the matrix polynomial equation (4.10), one can write

$$\lambda_j^2 e_{(j,1)} A + \lambda_j e_{(j,1)} B + e_{(j,1)} C = \underline{0} \tag{4.28}$$

Postmultiplying both side of the equation by $\underline{1}$ and then by using Property 4.2 one gets

$$\lambda_j^2 e_{(j,1)} A\underline{1}^T - \lambda_j e_{(j,1)} (A\underline{1}^T + C\underline{1}^T) + e_{(j,1)} C\underline{1}^T = 0 \text{ ,or ,}$$

$$\lambda_j(\lambda_j - 1) e_{(j,1)} A\underline{1}^T - (\lambda_j - 1) e_{(j,1)} C\underline{1}^T = 0 \text{ ,or ,}$$

$$(\lambda_j - 1)(\lambda_j e_{(j,1)} A\underline{1}^T - e_{(j,1)} C\underline{1}^T) = 0 \tag{4.29}$$

Thus, under the condition, λ_j is different than one, (4.27) holds. This is exactly what was to be proven.

(Q.E.D.)

Theorem 4.2:

$$e_{(j,k)}(\lambda_j A - C)\underline{1}^T = -e_{(j,k-1)} A\underline{1}^T \tag{4.30}$$

for all eigenvalues λ_j different from one and k=2,3,... $s(\lambda_j)$ where $s(\lambda_j)$ is the length of the generalized eigenvector cycle, or in other words, the size of the Jordan canonical form corresponding to the eigenvector, λ_j.

We refer to equation (4.30) as the Raw Balance Equations in the Component Form (RBECF). These equations are to be used for the generalized eigenvectors. For the first element of the cycle, which is the genuine eigenvector, BECF are used. But as one can observe for the case of generalized eigenvectors, the relation is a little bit more

complicated. Each expression depends on the previous element of the cycle. Yet, we are going to prove that this expression still yields a balance component by component. The impurity introduced by the left hand-side of (4.30) is actually essential for this occurrence since in the case of generalized eigenvectors, each term of the general solution starting with the coefficient corresponding to a certain generalized eigenvector also includes the previous eigenvectors of the same cycle. This fact is well manifested in the equation (4.19).

Proof:

We are going to prove this theorem using the induction method. First we have to show that the theorem is correct for k=2 which is the initial point for which the theorem is valid. The defining equation for the second eigenvector of a cycle is (see Lancester, Gohlberg, and Berman, 1982)

$$e_{(j,1)}\frac{dL(\lambda)}{d\lambda}+e_{(j,2)}L(\lambda)=\underline{0} \tag{4.31}$$

substituting $L(\lambda)$ from (4.11) one obtains

$$e_{(j,1)}(2\lambda_j A+B)+e_{(j,2)}(\lambda_j^2 A+\lambda_j B+C)=\underline{0} \tag{4.32}$$

Postmultiplying this equation with the $\underline{1}^T$ vector and then making use of Property 4.1 yields

$$e_{(j,1)}(2\lambda_j A-A-C)\underline{1}^T+e_{(j,2)}(\lambda_j^2 A-\lambda_j(A+C)+C)\underline{1}^T=0$$

Since the eigenvalue, λ_j, is different from one by hypothesis, one can use Theorem 4.1 and obtain

$$(\lambda_j-1)\left[e_{(j,1)}A\underline{1}^T+e_{(j,2)}(\lambda_j A-C)\underline{1}^T\right]=0\text{, or,}$$

$$e_{(j,2)}(\lambda_j A - C)\underline{1}^T = -e_{(j,1)}A\underline{1}^T \tag{4.33}$$

Thereby it is shown that the Theorem 4.2 is correct for k=2.

Then in order to use the induction argument one has to assume that the theorem is valid up to k=x and prove that this one ensures the validness of the theorem for k=x+1, i.e.,

$$e_{(j,x)}(\lambda_j A - C)\underline{1}^T = -e_{(j,x-1)}A\underline{1}^T \tag{4.34}$$

The defining equation for the $(x+1)^{th}$ eigenvector of a cycle is (see Lancester, Gohlberg, and Berman, 1982)

$$\sum_{y=0}^{x}\frac{1}{y!}e_{(j,x+1-y)}\frac{d^y L(\lambda)}{d\lambda^y} = \underline{0} \tag{4.35}$$

One can expand this expression by substituting $L(\lambda)$ from (4.11). The following expression is valid $x+1\geq 3$. Yet, since our induction starts at x=2, we can use it freely for our purposes.

$$e_{(j,x-1)}(A) + e_{(j,x)}(2\lambda_j A + B) + e_{(j,x+1)}(\lambda_j^2 A + \lambda_j B + C) = \underline{0} \tag{4.36}$$

Postmultiplying this equation with the $\underline{1}^T$ vector and then making use of Property 4.1 yields

$$e_{(j,x-1)}A\underline{1}^T + e_{(j,x)}(2\lambda_j A - A - C)\underline{1}^T + e_{(j,x+1)}(\lambda_j^2 A - \lambda_j(A + C) + C)\underline{1}^T = \underline{0}$$

Applying the induction hypothesis, equation (4.34), one obtains

$$(\lambda_j - 1)\left(e_{(j,x)}A\underline{1}^T + e_{(j,x+1)}(\lambda_j A - C)\underline{1}^T\right) = 0\text{, or}$$

$$e_{(j,x+1)}(\lambda_j A - C)\underline{1}^T = -e_{(j,x)}A\underline{1}^T$$

Thereby, we have shown that Theorem 4.2 is valid for k=x+1 given that it is valid for k=x. This along with the validity at k=2 is enough for the induction argument. Thus by induction, Theorem 4.2 is proven.

(Q.E.D.)

Now we would like to show that the balance of flow equations that holds for the general solution, as expressed by Property 4.4, can also be formulated for each component of the solution that corresponds to a nonzero eigenvalue. Although the idea is similar, the expression will be somewhat different due to nontrivial dependence of the solution components on the counting set.

Here, the word *component* is used for each expression that corresponds to a term starting with a $w_{(j,k)}$ in the equation (4.19). Thus, each component is a part of the solution that corresponds to a given genuine or generalized eigenvector. At total there are 2n components. This property is important since it is essential for the freedom of setting each $w_{(j,k)}$ arbitrarily, or more correctly, independently from the internal equations.

Before proceeding to the proof of Theorem 4.3, we have to formally define the components. One should note that each component is an ordered collection of M vectors, one for each element of the counting space. Let $Com^{i}_{(i,j)}$ be the i^{th} element of the collection that is the component that corresponds to k^{th} generalized eigenvector of the j^{th} eigenvalue, defined as

$$Com^{i}_{(j,k)} = \sum_{l=\max(i-k+1,0)}^{i} \binom{i}{l} \lambda_j^l e_{(j,k-i+l)} \tag{4.37}$$

Then the component, $Com_{(i,j)}$ can be defined as

$$Com_{(j,k)} = \left(Com^1_{(j,k)}, Com^2_{(j,k)}, \ldots, Com^M_{(j,k)}\right) \tag{4.38}$$

by concatenating the elements of the collection according to the order.

This notation is also used for the eigenvalues at the infinity. In this case, they are defined as

$$Com_{(j,k)} = \left(\underline{0}^T, \underline{0}^T, \ldots, Com^M_{(j,k)}\right)$$

$$\text{where } Com^M_{(j,k)} = \sum_{l=\max(M-k+1,0)}^{M} \binom{M}{l} e_{(j,k-i+l)} \tag{4.39}$$

After defining the entire notation needed, the theorem can now be stated.

Theorem 4.3:

$$Com^i_{(j,k)} C\underline{1}^T = Com^{i+1}_{(j,k)} A\underline{1}^T \tag{4.40}$$

for all (j,k) pair corresponding to an generalized eigenvector for which $\lambda_j \neq 1$ and for i = 0, 1, 3, . . ., M-1

We refer to equation (4.40) as the Generalized Balance Equations in Component Form (GBECF). We can readily notice that when k=1, these equations are the same as (4.27) which we have stated as Theorem 4.1. Thus, Theorem 4.1 is a special case of the Theorem 4.3. But this theorem is correct also for the generalized eigenvectors. This fact justifies name selected for the theorem.

Before attempting the proof of this theorem one more step needs to be taken. Now we are going to present a Lemma that we are going to use in proving Theorem 4.3. Before presenting the lemma, let us define

$$Q_x = \sum_{l=x}^{i} \lambda_j^l e_{(j,k-i+l)} \left[\lambda_j \binom{i+1}{l+1} A\underline{1}^T - \binom{i}{l} C\underline{1}^T \right] \tag{4.41}$$

Lemma 4.1:

i. $i-k+1 \geq 0 \Rightarrow Com_{(j,k)}^{i+1} A\underline{1}^T - Com_{(j,k)}^{i} C\underline{1}^T = Q_{i-k+1}$

ii. $i-k+1 < 0 \Rightarrow Com_{(j,k)}^{i+1} A\underline{1}^T - Com_{(j,k)}^{i} C\underline{1}^T = Q_0 + e_{(j,k-i-1)} A\underline{1}^T$

Proof:

Both part of lemma 4.1 can be easily shown by substituting Equation (4.37) and by changing the summation variable.

Lemma 4.2:

$$Q_x = -\binom{i}{x} \lambda_j^x e_{(k+x-i-1)} A\underline{1}^T \tag{4.42}$$

for $\max(i-k+2,0) \leq x \leq i$

Proof:

We prove this theorem by induction. The induction starts at point x = i and ends at point x = i-k+2. Thus the induction is done by decreasing the variable, x, one by one. First one has to show that the theorem is valid at the initial point. By the definition of Q_x,

$$Q_i = \lambda_j^i e_{(j,k)} (\lambda_j A\underline{1}^T - C\underline{1}^T)$$

When Theorem 4.2 is applied, one obtains

$$Q_i = -\lambda_j^i e_{(k-1)} A\underline{1}^T \tag{4.43}$$

Thereby the theorem is valid at x = i.

In order to apply the induction method, we assume that the theorem is valid from $x = i$ to $x = z$. At this point, we have

$$Q_z = -\binom{i}{z} \lambda_j^z e_{(k+z-i-1)} A\underline{1}^T \tag{4.44}$$

We know that

$$Q_{z-1} = Q_z + \lambda_j^{z-1} e_{(j,k+z-i-1)} \left[\lambda_j \binom{i+1}{z} A\underline{1}^T - \binom{i}{z-1} C\underline{1}^T \right] \tag{4.45}$$

Moreover, by algebraic manipulation, one can easily show that

$$\binom{i+1}{z} = \binom{i}{z} + \binom{i}{z-1} \tag{4.46}$$

By substituting (4.44) and (4.46) in (4.45) one obtains

$$Q_{z-1} = \binom{i}{z-1} \lambda_j^{z-1} e_{(j,k+z-i-1)} \left[\lambda_j A\underline{1}^T - C\underline{1}^T \right]$$

Since k+z-i-1 ≥ 2 we can make use of the Theorem 4.2 and thereby reach

$$Q_{z-1} = -\binom{i}{z-1} \lambda_j^{z-1} e_{(k+z-i-2)} A\underline{1}^T$$

We observe that the theorem is valid for x = z-1. Thus the proof by induction is complete.

(Q.E.D.)

Proof of Theorem 4.3:

In order to prove Theorem 4.3 two separate cases need to be investigated.

i. Case of i-k+1 ≥ 0

By lemma 4.1, we know that $Com^{i+1}_{(j,k)} A\underline{1} - Com^{i}_{(j,k)} C\underline{1} = Q_{i-k+1}$. Moreover from lemma 4.2 we have

$$Q_{i-k+2} = -\binom{i}{k-2}\lambda_j^{i-k+2} e_{(j,1)} A\underline{1}^T \text{, or,}$$

$$Q_{i-k+1} = -\binom{i}{k-2}\lambda_j^{i-k+2} e_{(j,1)} A\underline{1}^T + \lambda_j^{i-k+1} e_{(j,1)}\left[\lambda_j \binom{i+1}{k-1} A\underline{1}^T - \binom{i}{k-1} C\underline{1}^T\right]$$

Using equation (4.46) one obtains

$$Q_{i-k+1} = \binom{i}{k-1}\lambda_j^{i-k+1} e_{(j,1)}\left[\lambda_j A\underline{1}^T - C\underline{1}^T\right]$$

Applying Theorem 4.1 yields $Q_{i-k+1} = 0$,or ,

$$Com^{i}_{(j,k)} C\underline{1}^T = Com^{i+1}_{(j,k)} A\underline{1}^T$$

Thereby the theorem is proven for the first case.

ii. Case of i-k+1 < 0

By lemma 4.1, we know that

$$Com^{i+1}_{(j,k)} A\underline{1}^T - Com^{i}_{(j,k)} C\underline{1}^T = Q_0 + e_{(j,k-i-1)} A\underline{1}^T .$$

Moreover from lemma 4.2 we have $Q_0 = -e_{(j,k-i-1)} A\underline{1}^T$. This translates to $Com^{i}_{(j,k)} C\underline{1}^T = Com^{i+1}_{(j,k)} A\underline{1}^T$. Thereby the theorem is proven for the second case and the proof is complete.

(Q.E.D.)

Theorem 4.4:

There is one Jordan block that corresponds to the roots of the characteristic equation at one. That is, all roots at one give rise to a single cycle of generalized eigenvectors.

Proof:

If there existed more than one Jordan block that corresponds to the eigenvector, $\lambda_j = 1$, then the nullspace of $(A+B+C)^T$ would be greater than two. That is, A+B+C would be reducible. Since this would contradict Assumption 4.2, the theorem is proven by contradiction.

Theorem 4.5:

The eigenvalue, $\lambda_j = 1$, has multiplicity greater than two if and only if $e_{(J,1)} A \underline{1}^T = e_{(J,1)} C \underline{1}^T$.

We refer to Theorem 4.5 as the Non-Balance at Unity Theorem (NBUT). This is the equivalent of BECF of Theorem 4.1 for the case of the eigenvalue at one. Yet, we observe that when the multiplicity of the eigenvalue at one is one, there is no balance of flow in component form for the *component* that corresponds to this eigenvalue.

Proof:

We already showed that there is a single generalized eigenvector cycle corresponding to the eigenvalue at one. This means that the multiplicity of the eigenvalue at one is equal to the number of the elements of the mentioned cycle.

Now let us assume that the multiplicity of the eigenvalue at one is greater than one. Consequently, we know that equation (4.32), the defining equation for the second generalized eigenvector, is valid for this eigenvalue, i.e.,

$$e_{(J,1)}(2A+B)+e_{(J,2)}(A+B+C)=\underline{0} \tag{4.47}$$

Postmultiplying both sides of the equation with $\underline{1}^T$ and using Property 4.2 yields

$$e_{(J,1)}A\underline{1}^T=e_{(J,1)}C\underline{1}^T \tag{4.48}$$

Thus the theorem is shown in forward direction.

The proof in the reverse direction is a little more involved. For this we have to assume that $e_{(J,1)}A\underline{1}=e_{(J,1)}C\underline{1}$ and show that there is always some vector $e_{(J,2)}$ that would satisfy the defining equation for the second eigenvector which happens to be equation (4.47). In other words we have to show that we can always find a vector that is transformed to $-e_{(J,1)}(2A+B)$ by the transformation (A+B+C).

We can show that $-e_{(J,1)}(2A+B)$ is orthogonal to $\underline{1}$ by multiplying the two vectors and making use of Property 4.2 and equation (4.48) which is in our hypothesis, i.e.,

$$-e_{(J,1)}(2A+B)\underline{1}^T=-e_{(J,1)}(A-C)\underline{1}^T=0$$

At this point we are going to present some properties of the transformation (A+B+C). It is known that the rank of (A+B+C) is n-1 where n yields the dimension of

(A+B+C) by the expression nxn. This is due to Assumption 4.2 that states that (A+B+C) is an irreducible transition rate matrix. Consequently, S_R, the row space of (A+B+C), is n-1 dimensional. Moreover for any vector x, one can write

$$x(A+B+C)\underline{1}^T = 0$$

which means $\underline{1}$ is also orthogonal to the row space of (A+B+C). Thus a direct sum of span of $\underline{1}$ and S_R would be the entire space of n dimensional row vectors.

We can now conclude that $-e_{(j,1)}(2A+B)$ that is orthogonal to $\underline{1}$ has to be an element of S_R, the row space of (A+B+C). This means that we can always find a vector $e_{(j,2)}$ satisfying equation (4.47). Thus the multiplicity of the eigenvalue at one has to be greater than two.

(Q.E.D.)

Theorem 4.6:

For m≥3, the eigenvalue, $\lambda_j = 1$, has multiplicity greater than m if and only if $e_{(j,1)}A\underline{1}^T = e_{(j,1)}C\underline{1}^T$ and $e_{(j,k)}(A-C)\underline{1}^T = -e_{(j,k-1)}(A)\underline{1}^T$ for $2 \le k \le m-1$.

We refer to Theorem 4.6 as the Deficient Raw Balance at Unity Theorem (DRBUT). This is the equivalent of RBECF of Theorem 4.2 for the case of the eigenvalue at one. Yet, we observe for this case, the equation that would cause the balance of flow in component form for the *components* corresponding to the last element of the cycle of eigenvectors that belongs to the eigenvector at one, never holds. For the previous elements of the same cycle, equalities, which will cause the balance of flow in

component form for the *components* corresponding to these elements, hold. The equivalance between the balance of flow equations and these equalities will be shown by Theorem 4.7.

Proof:

The proof of the theorem will be based on induction. First we will show that the theorem is correct for m = 3. One should note that the induction argument should be used to prove a double implication.

Let us assume that the multiplicity of the eigenvalue at one is greater than three. Consequently, we can use Theorem 4.5 that asserts that $e_{(j,1)}A\underline{1} = e_{(j,1)}C\underline{1}$. Moreover we can use equation (4.36), the defining equation for $(x+1)^{th}$ eigenvector of the cycle, for the third eigenvector, and obtain

$$e_{(j,1)}(A) + e_{(j,2)}(2A + B) + e_{(j,3)}(A + B + C) = \underline{0} \tag{4.49}$$

Postmultiplying both sides of the equation with $\underline{1}^T$ and using Property 4.2 yields

$$e_{(j,2)}(A - C)\underline{1}^T = -e_{(j,1)}(A)\underline{1}^T \tag{4.50}$$

Thus, the theorem is proven in forward direction for m = 3.

Now we present the proof in the reverse direction for this case. For this endeavor, we have to assume that equations (4.50) and (4.48) hold and show that there is always some vector $e_{(j,3)}$ that would satisfy the defining equation for the third eigenvector which happens to be equation (4.49). In other words we have to show that we can always find a vector that is transformed to $-\left(e_{(j,1)}(A) + e_{(j,2)}(2A + B)\right)$ by the transformation (A+B+C).

We can show that $-\left(e_{(J,1)}(A)+e_{(J,2)}(2A+B)\right)$ is orthogonal to $\underline{1}$ by multiplying the two vectors and making use of Property 4.2 and equation (4.50) which is in our hypothesis, i.e.,

$$-\left(e_{(J,1)}(A)+e_{(J,2)}(2A+B)\right)\underline{1}^T=-\left(e_{(J,1)}(A)+e_{(J,2)}(A-C)\right)\underline{1}^T=0$$

In the proof of the Theorem 4.5 we had shown that the entire space of n dimensional row vectors can be expressed as a direct sum of the span of $\underline{1}$ and S_R, the row space of (A+B+C). Thereby we can conclude that $-\left(e_{(J,1)}(A)+e_{(J,2)}(2A+B)\right)$ which is orthogonal to $\underline{1}$ has to be an element of S_R. This means that we can always find a vector $e_{(J,3)}$ satisfying equation (4.49). Thus the multiplicity of the eigenvalue at one has to be greater than two, which means that the theorem is valid in reverse direction for m=3.

Now we assume that the theorem is correct for $m=x$, in order to apply the inductive method. We want to prove the theorem in forward direction for $m=x+1$. If the eigenvalue at one has multiplicity greater than $x+1$, then, since the multiplicity is greater than x, by the inductive hypothesis, we have $e_{(J,1)}A\underline{1}=e_{(J,1)}C\underline{1}$ and $e_{(J,k)}(A-C)\underline{1}^T=-e_{(J,k-1)}(A)\underline{1}^T$ for $2\le k\le x\text{-}1$. Moreover, since the multiplicity is greater than $x+1$ the defining equation for $(x+1)^{th}$ eigenvector of the cycle, equation (4.36), also holds. Postmultiplying both sides of the equation with $\underline{1}^T$ and using Property 4.2 yields

$$e_{(J,x)}(A-C)\underline{1}^T=-e_{(J,x-1)}(A)\underline{1}^T \tag{4.51}$$

Thus, the theorem is proven in forward direction for m = x+1.

The proof in the reverse direction for m=x+1 is similar to the proof in the case of m=3. We show that we can always find a vector that is transformed to

$-\left(e_{(j,s-1)}(A)+e_{(j,s)}(2A+B)\right)$ by the transformation (A+B+C). We do this with exactly the same methodology that has been applied to the case of m=3. Thus the proof of Theorem 4.6 by induction is completed.

(Q.E.D.)

Theorem 4.7:

For $\lambda_j = 1$, the following statements are true:

$$Com^{i}_{(j,k)}C\underline{1}^{T} = Com^{i+1}_{(j,k)}A\underline{1}^{T} \text{ for } i = 0,1,2,\ldots,M-1 \text{ and } k = 1,2,\ldots,s(\lambda_j)-1$$

$$Com^{i}_{(j,s(\lambda_j))}C\underline{1}^{T} \neq Com^{i+1}_{(j,s(\lambda_j))}A\underline{1}^{T} \text{ for } i = 0,1,2,\ldots,M-1$$

We refer to Theorem 4.7 as the Deficient Generalized Balance at Unity theorem (DGBUT). This is the equivalent of GBECF of Theorem 4.3 for the case of the eigenvalue at one. Yet, we observe for this case, there is no balance of flow in component form for the *components* corresponding to the last element of the cycle of eigenvectors that belongs to the eigenvector at one. For the previous elements of the same cycle, the balance of flow in component form for the components corresponding to these elements exists. One should notice that for $s(\lambda_j) = 2$, Theorem 4.7 is equivalent to Theorem 4.4. Thus Theorem 4.7 is a generalization of the Theorem 4.4 for the generalized eigenvector cycles of any length. This fact justifies the name selected for the theorem.

Proof:

Note that up to $k = s(\lambda_j) - 1$, Theorem 4.6 states exactly that the results that were valid for eigenvalues that are not at one, are valid for the case of the eigenvalue at one. When $\lambda_j = 1$ is substituted in the equation (4.27) of the Theorem 4.1 and in the equation (4.30) of the Theorem 4.2, we obtain the expressions of the Theorem 4.6. This means that the proof of Theorem 4.3 which is based on the equations (4.27) and (4.30) would also hold for the case of the eigenvalue at one up to $k = s(\lambda_j)-1$, the component corresponding to the eigenvector preceding the last eigenvector of the cycle. Since this result of Theorem 4.3 is identical with part i of this theorem, part i of the theorem is shown by following the steps of the proof of Theorem 4.3. These steps are not replicated here since the isomorphism is obvious at this point.

We now investigate the case of $k = s(\lambda_j)$. We can observe that we should have

$$e_{(j,s(\lambda_j))}(A-C)\underline{1}^T \neq -e_{(j,s(\lambda_j)-1)}(A)\underline{1}^T , \tag{4.51}$$

because if this were not the case, by Theorem 4.6 the multiplicity of the eigenvalue at one would have to be greater than $s(\lambda_j)$. Since this would be a contradiction, we should have equation (4.51). Yet, we know that if we had

$$e_{(j,s(\lambda_j))}(A-C)\underline{1}^T = -e_{(j,s(\lambda_j)-1)}(A)\underline{1}^T , \tag{4.52}$$

by the argument that we have done for the proof of part i, we could show that

$$Com^{i+1}_{(j,s(\lambda_j))}A\underline{1}^T - Com^{i}_{(j,s(\lambda_j))}C\underline{1}^T = 0 \tag{4.53}$$

In the proof of this identity, $e_{(J,s(\lambda_j))}(A-C)\underline{1}^T$ would be the first element of the summation that would add up to zero if (4.52) were to hold. Using this argument, we can observe that

$$Com^{i+1}_{(J,s(\lambda_j))}A\underline{1}^T - Com^{i}_{(J,s(\lambda_j))}C\underline{1}^T = e_{(J,s(\lambda_j))}(A-C)\underline{1}^T + e_{(J,s(\lambda_j)-1)}A\underline{1}^T \qquad (4.54)$$

By equation (4.51) the right-hand-side of equation (4.54) is different from zero. Thus we have shown that equation (4.54) does not hold. And this completes the proof of part ii.

(Q.E.D.)

Theorem 4.8:

The coefficient in the general solution, corresponding to the last element of the eigenvector cycle that corresponds to, $\lambda_j = 1$, is always zero.

Proof:

The equation (4.19) states the i^{th} element of the solution in an explicit form. Using the notation defined by equation (4.36), we can rewrite it in a more concise form

$$g_i = \sum_{j=1}^{\rho}\sum_{k=1}^{s(\lambda_j)} w_{(J,k)}Com^{i}_{(J,k)} + \delta(M,i)\sum_{j=\rho}^{\rho+\rho_\infty}\sum_{k=1}^{s(\lambda_j)} w_{(J,k)}Com^{M}_{(J,k)} \qquad (4.55)$$

Now we can substitute this expression into equation (4.24) whose validity is stated by Property 4.4 and obtain

$$\sum_{j=1}^{\rho}\sum_{k=1}^{s(\lambda_j)} w_{(J,k)}Com^{i}_{(J,k)}C\underline{1}^T + \delta(M,i)\sum_{j=\rho}^{\rho+\rho_\infty}\sum_{k=1}^{s(\lambda_j)} w_{(J,k)}Com^{M}_{(J,k)}C\underline{1}^T =$$
$$\sum_{j=1}^{\rho}\sum_{k=1}^{s(\lambda_j)} w_{(J,k)}Com^{i+1}_{(J,k)}A\underline{1}^T + \delta(M,i+1)\sum_{j=\rho}^{\rho+\rho_\infty}\sum_{k=1}^{s(\lambda_j)} w_{(J,k)}Com^{M}_{(J,k)}A\underline{1}^T$$

This expression is valid for i = 0, 1, 3, . . ., M-1. When Theorem 4.3 is applied, all the terms corresponding to eigenvalues different than one cancel out. Thereby, we are left just with the terms belonging to eigenvalue one, which incorporate only a single cycle of eigenvectors.

$$\sum_{k=1}^{s(\lambda_j)} w_{(j,k)} Com^{i}_{(j,k)} C\underline{1}^{T} = \sum_{k=1}^{s(\lambda_j)} w_{(j,k)} Com^{i+1}_{(j,k)} A\underline{1}^{T} \tag{4.56}$$

Applying Theorem 4.7, we observe that all terms that correspond to $k = 1,2,\ldots, s(\lambda_j) - 1$ cancel out. Thereby equation (4.56) is reduced to

$$w_{(j,s(\lambda_j))} \left(Com^{i+1}_{(j,s(\lambda_j))} A\underline{1}^{T} - Com^{i}_{(j,s(\lambda_j))} C\underline{1}^{T} \right) = 0$$

Part ii of Theorem 4.7 states that $Com^{i}_{(j,s(\lambda_j))} C\underline{1}^{T} \neq Com^{i+1}_{(j,s(\lambda_j))} A\underline{1}^{T}$. Thus $w_{(j,s(\lambda_j))} = 0$.

(Q.E.D.)

CHAPTER 5
IMPLICATIONS OF THE SPECTRAL THEORY ON MODELS OF PRODUCTION LINES

5.1 Introduction

The platform we have presented in Chapter 4 consists of the matrix polynomial solution procedure for the QBDPs and the spectral properties of QBDPs, which are closely tied the matrix polynomial solution procedure. We have stated that any QBDP process could be represented by certain block matrices that are defined in (3.2). Thus, the spectral theory developed in Chapter 4 can be applied to get insights about any system that can be modeled with the QBDPs.

In this chapter, we will first present the general implications of the spectral theory. That means we will attempt to associate physical meanings to the quantities that arise during the matrix polynomial procedure, and to explain how the spectral theory can be used to get insights for any given model.

Then we are going to instantiate these ideas on models of production lines. The models we are going to employ in this pursuit are the ones we have derived all along Chapter 3 in order to give concrete examples of QBDPs. We will thus demonstrate how the spectral theory can be employed as a means of getting insights on given models of QBDPs, while contributing to the production line research. Our methodology will allow us to reach formal conclusions previously unattainable.

Finally, by generalizing the results we gathered from the production line models, we will seek to reach general results about the production lines and eventually about the QBDPs. This chapter will constitute the link between our theoretical results and industrial engineering concerns. It will also further our understanding of the basic systems that has been the subject of many studies.

5.2 General Implications

The quantities of central importance in the matrix polynomial procedure are the eigenvalues of the characteristic equation for a given QBDP. In Chapter 4, we have demonstrated that for each QBDP that is modeled appropriately, one could easily associate a characteristic equation. Although the term characteristic equation belongs to the mathematical concept being employed, it is also quite befitting from a modeling perspective. The roots of the characteristic equation, the eigenvalues, determine the behavior of the solution, thereby the steady-state characteristics of the model.

The solution for a given QBDP model is always a linear combination of the *components* that we have defined by (3.37) and (3.38). Each *component* actually corresponds to a generalized eigenvector of an eigenvalue of the system. The closed-form expression for a component includes the value of the eigenvalue and the elements of the cycle of generalized eigenvectors from the first element to the given eigenvector. The general solution for a given QBDP is given in (3.55).

When we examine this solution structure we see that any *component* is a solution candidate. Furthermore the spectral theory shows us that each of these *components* acts like the solution on their own. That means, the elements forming the solution all have the properties of the full solution. This is demonstrated by the fact that all the *components* have a balance of flow property within themselves. Thus, the balance of flow property for the full solution is not a property that manifests itself only at that level, it is the consequence of the fact that each of the elements that form the solution exhibit it on their own.

Each *component* consists of two building blocks: the eigenvalue and the eigenvectors–just one if the eigenvalue has a simple eigenvector corresponding to it, otherwise the elements of the generalized eigenvector cycle. Yet, the eigenvalue is the more crucial block since it determines the behavior of the *component*. If the eigenvalue's norm is greater than one, the *component* that corresponds to the eigenvalue will become more and more pronounced for larger elements of the counting space. Complementarily, if the eigenvalue's norm is smaller than one, the *component* that corresponds to the eigenvalue will become less and less pronounced for larger elements of the counting space. If the eigenvalue is at one, the component will have equal contribution all over the counting set. Furthermore, if the eigenvalue is complex, one would observe an oscillatory behavior in the *component* corresponding to it.

If we have the eigenvalues of the system at hand we can tell quite a bit about the possible behavior of the system. As in the control theory one could even try to make a root-locus diagram for the eigenvalues for design and sensitivity analysis purposes. If one

knows how the system parameters change the eigenvalues of the system, one can use this information to perturb the system towards a desired behavior.

Thus, the eigenvalues determine the conduct of the *components*, which are actually candidate solutions. One can compare them with modes that present themselves in a more established scientific field, electromagnetic theory. The modes are the possible electromagnetic waves that can exist in a given space, and boundary conditions determine what kind of combination of the modes would actually be observed in the space. Here, similarly, *components* are possible solutions of the system, and the boundary equations determine which combination of them is the final solution to the system yielding the steady-state probabilities.

One can imagine that by perturbing the boundary equations, it would be possible to change the weight of the modes and thereby to push the system in a desired direction. This would ipso facto amount to choosing the desirable modes of the system and making them more pronounced by changing the conduct of the system only at those states corresponding to the boundaries of the counting set.

Yet, one should be careful with the fact that although the modes are possible solutions in an independent fashion, they are still interrelated. For example, Theorem 4.8 states that the *component* corresponding to the last generalized eigenvector of the eigenvalue at one, has always zero weight, which means that it does not appear in the final solution. Thus, although this *component* is a candidate solution per se, it can never be a part of the solution.

Moreover, one can observe that certain eigenvalues work in groups. Numerical experimentation shows that in certain cases, certain elements of a *component* may be

negative, which would not make any sense if it were to be elected to be a solution in isolation. Yet, since it is always balanced with another *component* in an intrinsic fashion, whatever the boundary equations are, never a negative number appears in the final solution.

The intrinsic relation between the *components* is crucial because otherwise the solutions obtained through the matrix polynomial procedure could never be valid. Although it is not quite possible to grasp this interrelation in a theoretical way at this point with the tools we have presented, we are going to demonstrate it on actual examples in the next section of the chapter.

In this section, we have exposed the general implications of the spectral theory that we have developed in Chapter 4. That means we have discussed how the theory can be exploited for a better understanding of any given model of QBDP. But whenever something is put in such general terms it always carries the curse of being vague. This curse will be remedied with the application of the theory on concrete models that will be presented in the next section.

5.3 Application of the Spectral Theory to Specific Models of Production Lines

Although QBDPs manifest themselves in the stochastic modeling of many systems, in this section, we are going to restrict ourselves only to those models that we have established in Section 3.3 of the dissertation. The QBDP formulations obtained in Section 3.3 will be directly utilized in our analysis. Thereby, the reader should consult

that section whenever a clarification about the model and the underlying assumptions is needed.

The models of concern pertain to production lines. Thus, we aspire to demonstrate how the spectral analysis can be utilized to bring about fresh ideas in an already well established field. However, we would like to point out that the methodology applied on the given instances of QBDPs is readily extendible to any other QBDP. Thus, the prospects of this effort are higher than it may initially appear.

Model 1: Exponential Server With Limited Buffer:

This is the quintessential M/M/1/M model, probably the most investigated model of the whole queuing literature. Every treaty in the subject starts by analyzing this model. So as not to contradict this custom, we set ourselves to investigate it in the light of the matrix polynomial solution procedure.

The characteristic equation for the model is:

$$\det(L(x)) = \mu x^2 - (\lambda + \mu)x + \lambda \tag{5.1}$$

One needs to find eigenvalues of the system, the roots of the characteristic equation

$$\Delta = (\lambda + \mu)^2 - 4\lambda\mu = (\lambda - \mu)^2 \tag{5.2}$$

$$x_1 = 1 \ , \ x_2 = \frac{\lambda}{\mu} \tag{5.3}$$

The fact that the discriminant of this second-degree equation is always zero, ensures that the roots are always real. The two roots can be distinct or non-distinct according to the choice of parameters. Wherever $\lambda = \mu$, the roots become non-distinct,

and one obtains two eigenvalues at one, for the system. Thus the case of equality and of inequality of the two parameters of the model have to be investigated separately.

i. Case of $\lambda = \mu$

For this case, we have one eigenvalue for the model. And that eigenvalue is at one. We perceive that when $\lambda = \mu$, the system is balanced, which means that the average incoming arrival rate is equal to the average processing rate of the server. And we would like to postulate that whenever this equality occurs, we expect an additional multiplicity for the eigenvalue at one. Indeed, for this basic model our postulate is valid. The component that belongs to this addition to the generalized eigenvector cycle at one will be constant all along the different buffer levels. That means it will not be biased towards the smaller of the larger buffer levels. Consequently, this component will be distributed in a balanced fashion along the counting set. Thus, it is not difficult to link the additional eigenvalue at one to the balance of the system.

Theorem 4.6 suggests that we should have $e_{(1,1)}A\underline{1}^T = e_{(1,1)}C\underline{1}^T$ for the additional multiplicity at one. If one actually scrutinizes this condition, it requires the balance of rate of passage from a buffer level group to a higher buffer level group, with the rate of passage from a buffer level group to a lower buffer level group. Thus, it relates to the balance of the system. In fact, this condition for the case of this model is the same with our condition, $\lambda = \mu$.

Since for this model the submatrices are scalars the cycle of generalized eigenvectors at one consists of two constants. By Theorem 3.8, the *component* that

corresponds to the second element of the cycle does not appear on the solution. Thus, the solution is

$$g_i = k \quad for\ 1 \le i \le M \tag{5.4}$$

Upon the application of the normalizing condition, the solution as expected becomes

$$g_i = \frac{1}{M+1} \quad for\ 1 \le i \le M \tag{5.5}$$

The solution is constant all over the buffer levels. We can observe that the system is fully balanced. The probability of being in any state is the same.

ii. Case of $\lambda \neq \mu$

For this case, we have two distinct eigenvalues for the model. One of these eigenvalues is at one and the other is at $\frac{\lambda}{\mu}$. The second eigenvalue is known in queuing literature as the traffic intensity or the utilization of the queue and is noted by ρ. By the Theorem 4.8, the *component* that corresponds to the eigenvalue at one, does not appear in the solution. Thence, only the second eigenvector is in the solution. Moreover, since the submatrices are, for this case, scalars, the eigenvector that corresponds to this eigenvalue is a constant. Thus, the solution is

$$g_i = k\rho^i \quad for\ 1 \le i \le M \tag{5.6}$$

Upon the application of the normalizing condition, the solution becomes:

$$g_i = \left(\frac{1-\rho}{1-\rho^{M+1}}\right)\rho^i \quad for\ 1 \le i \le M \tag{5.7}$$

This is the typical M/M/1/M result that we would expect. Notice that we do not need to impose any restriction on the parameter space, since even with eigenvalues with norm larger than one, the solution is convergent. This is due to the fact that the finite state-space Markov processes, given that periodicity does not exist, the process is always ergodic.

At this stage, we have demonstrated that our results are in full accordance with the previously known results as one would expect. But we have not reached any new conclusion except a justification for the relation between the balance of the system and the eigenvalues at one. The simplicity of the analyzed system prevented us from fully evincing the power of the spectral theory to unearth new results.

Model 2: Exponential Server with Breakdown and Repair, Limited Buffer:

This model is a basic extension of the M/M/1/M model. The breakdown and repair is incorporated to it. Yet even with this much addition, the conduct of our model becomes much more complicated. Because of it, this is a good model to demonstrate the power of our methodology. Therefore we will specially emphasize the analysis of this model. Some of the concepts that will be introduced here, will be later extended to the following models.

The matrix polynomial that corresponds to this QBDP is

$$L(x)=\begin{bmatrix} \mu x^2-(\alpha+\lambda+\mu)x+\lambda & \alpha x \\ \beta x & -(\beta+\lambda)x+\lambda \end{bmatrix} \tag{5.8}$$

The characteristic equation can be easily derived from the matrix polynomial

$$\det(L(x)) = -\mu(\lambda+\beta)x^3 + (2\lambda\mu + \alpha\lambda + \lambda^2 + \beta\lambda + \beta\mu)x - (\alpha\lambda + \lambda\mu + \beta\lambda + 2\lambda^2)x + \lambda^2 \tag{5.9}$$

According to the theory, one would expect to have four eigenvalues corresponding to a QBDP with submatrices of dimension 2x2. Yet the characteristic equation being a polynomial of degree three, can only provide us with three eigenvalues. Yet, since the nullspace of A is of dimension one, we have another eigenvalue at the infinity.

Another root that is always known is one by Property 4.3. This fact is used to check the validity of our derivation, and to ease the retrieval of the other two finite roots. Once the det(L(x)) of (5.9) is divided to (x-1), one obtains a second degree polynomial that embeds the critical two eigenvalues of the system:

$$\frac{\det(L(x))}{x-1} = -\mu(\lambda+\beta)x^2 + \lambda(\mu+\alpha+\lambda+\beta)x - \lambda^2 \tag{5.10}$$

The two eigenvalues embedded in the equation are critical since they are dependent on the system's parameters. Thereby, they elucidate the relation between the system parameters and the nature of the steady-state probabilities. In order to find and classify the nature of the eigenvalues one should first investigate the discriminant of the second-degree equation of (5.10).

$$\Delta = \lambda^2\left((\alpha+\beta+\lambda+\mu)^2 - 4\mu(\beta+\lambda)\right) \tag{5.11}$$

It is possible to write the discriminant in another form:

$$\Delta = \lambda^2\left((\alpha+\beta+\lambda-\mu)^2 + 4\mu\alpha\right) \tag{5.12}$$

This form is useful since it shows that the discriminant is always positive. Since it is known the parameters are always positive, the expression $\left((\alpha+\beta+\lambda-\mu)^2 + 4\mu\alpha\right)$ is always positive. Thereby, the discriminant is always positive. This is important since it

shows that the eigenvalues of the system are all real. That is no complex eigenvalue, which could cause an oscillating set of steady-state probabilities.

At this stage one can express in closed-form all the eigenvalues of the system:

$$x_1 = \frac{\lambda}{2\mu}\left(\left(1+\frac{\alpha+\mu}{\beta+\lambda}\right)-\sqrt{\left(1+\frac{\alpha+\mu}{\beta+\lambda}\right)^2-\frac{4\mu}{\beta+\lambda}}\right) \tag{5.13}$$

$$x_2 = \frac{\lambda}{2\mu}\left(\left(1+\frac{\alpha+\mu}{\beta+\lambda}\right)+\sqrt{\left(1+\frac{\alpha+\mu}{\beta+\lambda}\right)^2-\frac{4\mu}{\beta+\lambda}}\right) \tag{5.14}$$

$x_3 = 1 \quad , \; x_4 = \infty$

As one can see, the first two eigenvalues are more critical for the understanding of the system, since they depend on the system parameters. The eigenvalue at one will not affect the solution by Theorem 4.8. At this point, one needs to address issues such as the dependence of two eigenvalues (5.13 and 5.14) to the system parameters and their range of values.

One thing that can be observed right away is that the square root of the discriminant (5.11) is always smaller than $\left(1+\frac{\alpha+\mu}{\beta+\lambda}\right)$; thence, both eigenvalues (5.13 and 5.14) are always positive. If not, oscillatory behavior could have been observed in the solution.

All the parameters that we make use of here have the unit, 1/[Time Unit]. Therefore their actual values depend on the choice of time unit. Yet one can express the eigenvalues in terms of parameters that are independent of the choice of time unit, while decreasing the number of parameters by one. In Model 1, this is done by the introduction

of ρ, the traffic intensity. For this model, the same parameter can also be employed but we need to define two more parameters:

$\gamma = \frac{\alpha}{\lambda}$, Breakdown intensity.

$\delta = \frac{\beta}{\lambda}$, Repair intensity.

These definitions are useful from an engineering point of view, since it relates the eigenvalues of the system, to physically meaningful quantities. Thence, they make the relation between the system assumptions and the eigenvalues, which determine the conduct of the system, more visible.

When (5.13) and (5.14) are expressed in terms of the new parameters, one gets

$$x_1 = \frac{\rho}{2}\left(\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right) - \sqrt{\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)^2 - \frac{4}{\rho(1+\delta)}}\right) \tag{5.15}$$

$$x_2 = \frac{\rho}{2}\left(\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right) + \sqrt{\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)^2 - \frac{4}{\rho(1+\delta)}}\right) \tag{5.16}$$

At this stage, it is useful to draw these two eigenvalues as a function ρ, for certain physically meaningful choice of parameters γ and δ. This would enables us to understand how the systems behavior changes as the traffic intensity changes. It is reasonable to set $\delta = .1$ and $\gamma = .01$, since normally there is an order of magnitude of difference between the arrival rate and the repair rate, and between the repair rate and the breakdown rate. Otherwise, the machine would be implausibly inefficient.

Figure 5.1 clearly shows what happens as the traffic intensity changes. The first eigenvalue starts from a point under the unity level, and after crossing the unity level it

tends to infinity. Whereas the second eigenvalue starts from zero and asymptotically approaches a level below the unity. The second eigenvalue never crosses the unity level. Both of them are monotonically increasing functions. To further support these observations we present the eigenvalue functions for different system parameters, γ and δ, in Figure 5.2.

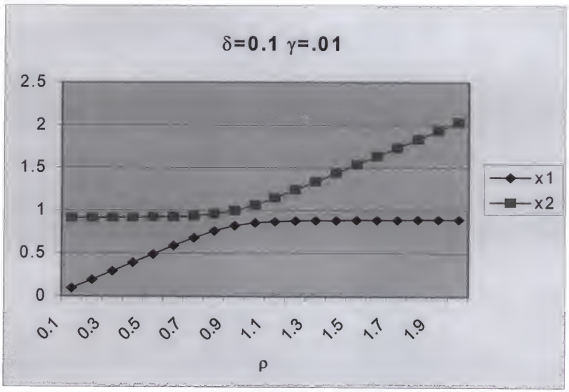

Figure 5.1 Eigenvalues drawn as a function of ρ, for $\delta = .1$ and $\gamma = .01$.

Having observed the eigenvalues' conduct for different set-ups in Figure 5.2, our initial remarks seem to be valid. At this stage, we attempt to formalize our inferences by the use of mathematical techniques. We first set to investigate the limiting behavior of

the eigenvalues–that means where they tend as the parameters take their limiting values, 0 and infinity.

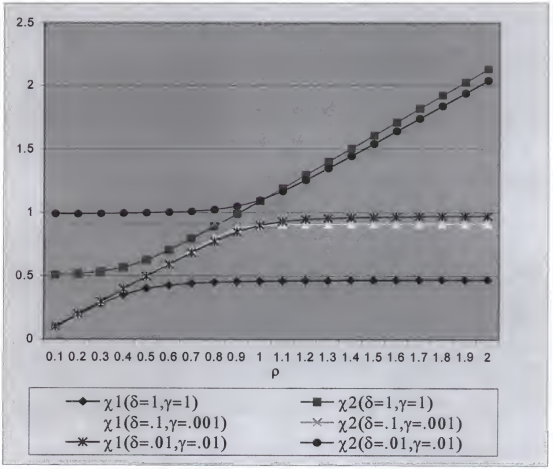

Figure 5.2 Eigenvalues drawn as a function of ρ, functions for different system parameters, γ and δ.

As ρ tends to zero from the right,

$$x_1 \to \frac{\rho}{2}\left(\left(1+\frac{1}{\rho(1+\delta)}\right)-\sqrt{\left(1+\frac{1}{\rho(1+\delta)}\right)^2-\frac{4}{\rho(1+\delta)}}\right)$$

$$x_1 \to \frac{\rho}{2}\left(\left(1+\frac{1}{\rho(1+\delta)}\right)-\sqrt{\left(1-\frac{1}{\rho(1+\delta)}\right)^2}\right)=\frac{\rho}{2}\left(\left(1+\frac{1}{\rho(1+\delta)}\right)-\left|1-\frac{1}{\rho(1+\delta)}\right|\right)$$

Since as ρ tends to zero $\frac{1}{\rho(1+\delta)} >> 1$, $\left|1-\frac{1}{\rho(1+\delta)}\right| < 0$, therefore $x_1 \to \rho \to 0$.

Thus, $\lim_{\rho \to 0^+} x_1 = 0$.

Using the same steps, we reach the conclusion $\lim_{\rho \to 0^+} x_2 = \frac{1}{1+\delta}$

As ρ tends to infinity,

$$x_1 \to \frac{\rho}{2}\left(\left(1+\frac{\rho\gamma}{\rho(1+\delta)}\right)-\sqrt{\left(1+\frac{\rho\gamma}{\rho(1+\delta)}\right)^2}\right) \to \infty.0, \text{ and}$$

$$x_2 \to \frac{\rho}{2}\left(\left(1+\frac{\rho\gamma}{\rho(1+\delta)}\right)+\sqrt{\left(1+\frac{\rho\gamma}{\rho(1+\delta)}\right)^2}\right) \to \rho\left(1+\frac{\gamma}{(1+\delta)}\right) \to \infty$$

Thus, we find that $\lim_{\rho \to \infty} x_2 = \infty$ (moreover it tends to infinity like the function, $\left(1+\frac{\gamma}{(1+\delta)}\right)\rho$), but the limit of $\lim_{\rho \to \infty} x_1$ is still undetermined after the given analysis.

That means some other analysis is needed. At this point, when L'Hopital's Theorem is applied, one obtains

$$\lim_{\rho \to \infty} \frac{1}{2} \frac{\left(\frac{-\rho^{-2}}{(1+\delta)}-\frac{1}{2}\left(\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)^2-\frac{4}{\rho(1+\delta)}\right)^{-\frac{1}{2}}\left(2\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)\frac{-\rho^{-2}}{(1+\delta)}+4\frac{\rho^{-2}}{(1+\delta)}\right)\right)}{-\rho^{-2}}$$

$$=\frac{1}{2}\left(\frac{1}{(1+\delta)}-\frac{1}{2}\left(1+\frac{\gamma}{(1+\delta)}\right)^{-1}\left(2\left(1+\frac{\gamma}{(1+\delta)}\right)\frac{1}{(1+\delta)}-\frac{4}{(1+\delta)}\right)\right)=\frac{1}{1+\delta+\gamma}$$

Thus, we have shown that $\lim_{\rho \to \infty} x_1 = \frac{1}{1+\delta+\gamma}$.

At this stage, we have been able to isolate two asymptotes for each eigenvalue. The first eigenvalue is asymptotic to the function $f(\rho) = \rho$ as it approaches to zero, and asymptotic to $f(\rho) = \frac{1}{1+\delta+\gamma}$ as it tends to infinity. Likewise, the second eigenvalue is asymptotic to $f(\rho) = \frac{1}{1+\delta}$ as it approaches to zero, and asymptotic to $f(\rho) = \left(1 + \frac{\gamma}{1+\delta}\right)\rho$ as it tends to infinity.

At this stage, to show that the eigenvalues are monotonically increasing functions of the traffic rate, ρ, we examine the eigenvalues' derivatives with respect to ρ:

$$\frac{\partial x_1}{\partial \rho} = \frac{1}{2}\left(\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right) - \sqrt{\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)^2 - \frac{4}{\rho(1+\delta)}}\right) + \frac{\rho}{2}\left(\frac{-\rho^{-2}}{1+\delta} - \frac{1}{2}\left(\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)^2 - \frac{4}{\rho(1+\delta)}\right)^{-\frac{1}{2}}\left(2\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)\frac{-\rho^{-2}}{1+\delta} + \frac{4\rho^{-2}}{1+\delta}\right)\right)$$

$$\frac{\partial x_2}{\partial \rho} = \frac{1}{2}\left(\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right) + \sqrt{\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)^2 - \frac{4}{\rho(1+\delta)}}\right) + \frac{\rho}{2}\left(\frac{-\rho^{-2}}{1+\delta} + \frac{1}{2}\left(\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)^2 - \frac{4}{\rho(1+\delta)}\right)^{-\frac{1}{2}}\left(2\left(1+\frac{1+\rho\gamma}{\rho(1+\delta)}\right)\frac{-\rho^{-2}}{1+\delta} + \frac{4\rho^{-2}}{1+\delta}\right)\right)$$

When the roots of both of the derivatives is investigated using a symbolic mathematical tool called *Maple*, one finds out both of them have no roots. Thence, both derivatives are always positive. This means that the derivatives are always strictly

positive. Thus, both eigenvalues are monotonically increasing functions with respect to ρ.

At this point, one can conclude that the both of the components are more biased towards the smaller buffer levels when the traffic intensity is relatively small. As the traffic intensity increases, this bias becomes less and less evident. At one point, the second eigenvalue transcends the unity. After this point, the second component is biased towards the larger buffer sizes. Yet, the first component is always more prominent at smaller buffer levels.

The point at which the second eigenvalue crosses the unity is also of utmost importance. At that point the second eigenvalue is at unity, which means that the component corresponding to it is balanced over the buffer levels. We have already postulated that an additional eigenvalue at one always relates to the balance of the system. Using *Maple* one can obtain the roots of the equation, $x_2 = 1$. The solution to this equation is

$$\rho = \frac{\delta}{\delta + \gamma} \tag{3.17}$$

As we can easily show, the effective average production rate of the system–adjusted for breakdowns–is $\mu \frac{\beta}{\alpha + \beta}$. The balance occurs when the average production rate is the same with the average arrival rate, λ. Thus for a balanced system one should have

$$\lambda = \mu \frac{\beta}{\alpha + \beta} \tag{3.18}$$

This result is quite remarkable, since it is obvious that (3.18) is equivalent to (3.17). Thus, the balance conditions one derives for a balanced system through solely physical arguments, gives us an eigenvalue at one. This shows lucidly that the eigenvalues at one relate to the balance of system. And for more complicated problems, one can always check numerically if there are eigenvalues close enough to one.

One should observe that from a perspective of averages, only the proportion of parameters γ and δ matters. Yet, for the functional forms of the eigenvalues we see that the terms appear in such forms that they cannot be reduced to a proportion. Thus, even when the proportion of length of the broken periods to length of working period is kept constant, the system behavior changes. One needs to investigate the extent of this change, by examining at the eigenvalues for different system parameters having a constant proportion.

Upon the examination of the Figure 5.3, one should induce that the values taken by the parameters γ and δ are of concern in an independent fashion. Thence, the proportion of broken and busy cycles does not yield sufficient information to determine the system behavior. The length of the duty cycle relative to the interarrival cycle is also critical. This is physically due to the fact that the integer number of parts processed in a duty cycle is not fully scalable with the cycle length.

Up to this point, we have delved into the analysis of the eigenvectors as a function of the traffic intensity. Although the variation of the system behavior with the other two intensities is also of concern, their importance is relatively less. This is due to the fact that the traffic intensity of the system exists for all queuing systems, and therefore an

analysis based on traffic intensity can be comparable with the analyses of systems with different characteristics.

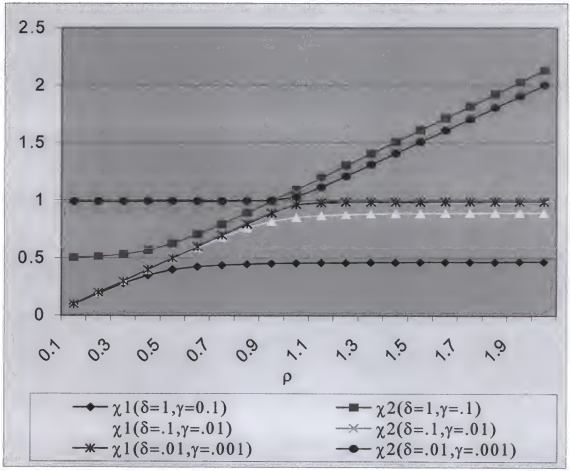

Figure 5.3 Eigenvalues drawn as function of ρ, while $\frac{\delta}{\gamma}=10$.

Still, we investigate the tendencies of the eigenvalues for the limiting values of the other parameters, γ and δ:

$$\lim_{\delta\to 0^+} x_1 = \frac{\rho}{2}\left(\left(1+\frac{1}{\rho}+\gamma\right)-\sqrt{\left(1+\frac{1}{\rho}+\gamma\right)^2-\frac{4}{\rho}}\right)$$

$$\lim_{\delta \to 0^+} x_2 = \frac{\rho}{2}\left(\left(1 + \frac{1}{\rho} + \gamma\right) + \sqrt{\left(1 + \frac{1}{\rho} + \gamma\right)^2 - \frac{4}{\rho}}\right)$$

$$\lim_{\delta \to \infty} x_1 = 0\,,\ \lim_{\delta \to \infty} x_2 = \rho$$

$$\lim_{\gamma \to 0^+} x_1 = \rho\,,\ \lim_{\gamma \to 0^+} x_2 = \frac{1}{1+\delta}$$

$$\lim_{\gamma \to \infty} x_1 = 0\,,\ \lim_{\gamma \to \infty} x_2 = \infty$$

As the repair intensity tends to zero, the server stands for longer times in the repair states. That means less production occurs, and thereby more buildup is observed in the buffers, given that the system is fed by sufficient arrivals.

As the repair intensity tends to infinity, the server tends to spend less and less time during repairs. Thus the system becomes more and more like a typical M/M/1 system with no repair states. This one is visible from the fact that the only finite eigenvalue in the limiting case is ρ, which is the eigenvalue for an M/M/1 system.

As the breakdown intensity tends to zero, once again the system tends to an M/M/1 system. Yet for this case, we have another non-finite eigenvalue that is independent of the traffic intensity.

As the breakdown intensity tends to infinity, one of the eigenvalues tends to zero and the other to infinity. That means the system has zero probability of being in buffer levels between 1 and M-1. All the probabilities are shared by states at buffer levels 0 and M. This is due to the fact that the processing of parts is not possible under such conditions.

After this rather thorough analysis of the eigenvalues, let us go through the full matrix polynomial procedure for a given system so that the procedure is elucidated and some of the occurrences are illustrated.

Let $\alpha = .01,\ \beta = .1,\ \lambda = 1,\ \mu = 1$, M=10 then

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} -2.01 & 0.01 \\ 0.1 & -1.1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} -1.01 & 0.01 \\ 0.1 & -1.1 \end{bmatrix}, B_M = \begin{bmatrix} -1.01 & 0.01 \\ 0.1 & -0.1 \end{bmatrix}$$

The characteristic equation for the system is:

$$\det(L(x)) = -1.1x^3 + 3.2x^2 - 3.11x + 1$$

Thence, the eigenvalues for the system are

$$x_1 = 0.8553, \quad x_2 = 1.0628, \quad x_3 = 1, \quad x_4 = \infty$$

We also need the eigenvectors corresponding to these eigenvalues in order to construct the general solution and the boundary conditions. These eigenvectors along with the corresponding components are

$$e_{(1,1)} = [-0.9897 \quad 0.1432], \ Com^{i}{}_{(1,1)} = (0.8853)^{i}[-0.9897 \quad 0.1432]$$

$$e_{(2,1)} = [0.9980 \quad 0.0627], \ Com^{i}{}_{(2,1)} = (1.0628)^{i}[0.9980 \quad 0.0627]$$

$$e_{(3,1)} = [0.9950 \quad 0.0995], \ Com^{i}{}_{(3,1)} = [0.9950 \quad 0.0995]$$

$$e_{(4,1)} = [0 \quad 1], \ Com_{(4,1)} = [0 \quad 1]$$

It is interesting to note that one of the entries in the eigenvector $e_{(1,1)}$ is negative. Thus, for any given choice of coefficient for the component corresponding to this

eigenvector, the contribution of the component for one of the dimensions will be negative. This one indeed shows that this component can never be a solution in isolation, although it still complies with the conservation of flow. Thus, in any physically meaningful solution it needs to be balanced with the component corresponding to the eigenvector $e_{(2,1)}$. Otherwise, the solution could have negative entries, which would contradict the very foundations of probability theory.

At this stage, one can write the boundary equations a matrix notation:

$$\begin{bmatrix} w_{(1,1)} & w_{(2,1)} & w_{(3,1)} & w_{(3,1)} \end{bmatrix} \begin{bmatrix} 0.1674 & -0.1674 & -0.0300 & 0.0300 \\ 0.0590 & -0.0590 & -0.1154 & 0.1154 \\ 0 & -0.0995 & 0 & 0.0995 \\ 0 & 0 & 0.1000 & -0.1000 \end{bmatrix} = \underline{0}$$

We see that the boundary equations form a homogeneous system. For a nontrivial solution to exist, the rank of this system should be at most three. But this is always the case for the boundary equations. Thus the solution to the system is actually the nullspace of the boundary matrix. For the presented system, the nullspace is

$$\begin{bmatrix} w_{(1,1)} & w_{(2,1)} & w_{(3,1)} & w_{(3,1)} \end{bmatrix} = \begin{bmatrix} -0.2365 & 0.6708 & 0 & 0.7029 \end{bmatrix}$$

Here we observe that the coefficient that corresponds to the first component is negative. We have already recognized that the corresponding component has two entries of the opposite signs. Thence, this only changes the place of the negative and the positive entry. The contributions due to the other eigenvalues are all positive. We also underline the fact that the coefficient of the eigenvalue at one is zero as predicted by Theorem 4.7.

After having obtained the weights of the components in the solution one can generate the solution. Then by normalizing the solution one can finally reach the steady-state probabilities for the model:

We see that after going through all the steps, although some components have negative contributions to the solution, the solution yields a set of steady-state probabilities that is sensible. That means the interference between the eigenvalue-eigenvector pairs somehow insures the positiveness of the entries in the solution.

Table 5.1 Steady-state probabilities for a given model of exponential server with breakdown and repair, limited buffer.

M	0	1	2	3	4	5	6	7	8	9	10
W	0.0851	0.0833	0.0821	0.0812	0.0808	0.0806	0.0808	0.0812	0.0818	0.0826	0.0836
R	0.0008	0.0013	0.0018	0.0022	0.0026	0.0030	0.0033	0.0036	0.0038	0.0041	0.0705

In order to better observe the delicate balance between the two components associated with the first two eigenvalues, one can check their contribution to the final solution. That means the inspection of the entries of $w_{(1,1)}Com_{(1,1)}$ and $w_{(2,1)}Com_{(2,1)}$ that corresponds to the repair states, since the contribution of $w_{(1,1)}Com_{(1,1)}$ is negative and it has to be balanced by $w_{(2,1)}Com_{(2,1)}$.

Table 5.2 Negative (1st row) and positive (2nd row) contributions to an entry of the solution by the first and the second components and their combination (3rd row)

0	1	2	3	4	5	6	7	8	9	10
-0.034	-0.029	-0.025	-0.021	-0.018	-0.016	-0.013	-0.011	-0.010	-0.008	-0.007
0.042	0.043	0.044	0.045	0.046	0.047	0.048	0.049	0.050	0.052	0.053
0.008	0.014	0.019	0.024	0.028	0.032	0.035	0.038	0.041	0.043	0.046

We see that both contributions are monotonically increasing functions of the buffer level. An interesting observation is that the positive contributions increase faster than the negative contributions as the buffer level increases. Thus, given that the component coefficients are set such that the positive contribution is greater than the negative contribution for the zero buffer level, the solution will be physically plausible.

Since it is the boundary equations, which determines the value of the coefficients, if one perturbs the boundary equations, the resulting coefficient set would be different. Yet, one would always obtain a coefficient set such that the positive contribution is greater than the negative contribution for the zero buffer level. Thus, one cannot set the coefficients as one desires by perturbing the boundary equations. The control that can be exerted in manipulating the coefficient set will always be limited.

For our example the proportion of the first coefficient to the second coefficient is

$$\frac{w_{(1,1)}}{w_{(2,1)}} = -0.3526.$$

We can derive a theoretical bound on it by using the eigenvector entries:

$$w_{(2,1)}(0.0627) + w_{(1,1)}(0.1432) \geq 0 \Rightarrow \frac{w_{(1,1)}}{w_{(2,1)}} \geq -0.4378 \qquad (5.15)$$

Another bound can be derived using the other entry of both eigenvectors:

$$w_{(2,1)}(0.9980) + w_{(1,1)}(-0.9897) \geq 0 \Rightarrow \frac{w_{(1,1)}}{w_{(2,1)}} \leq 1.0084 \qquad (5.16)$$

Thence, even when the boundary equations are manipulated the proportion of the coefficients should be in the range (-0.4378 , 1.0084).

At this stage, one should try to analyze how this proportion of coefficients varies when the boundary equations are perturbed. The only way to change the boundary equations without touching the components of the system, is to perturb the matrices B_0 and B_M. A good way doing it is to differentiate the breakdown and the repair rates at the boundary states. Then, the matrices can be written as

$$B_0 = \begin{bmatrix} -(\alpha_0 + \lambda) & \alpha_0 \\ \beta_0 & -(\beta_0 + \lambda) \end{bmatrix}, B_M = \begin{bmatrix} -(\alpha_M + \mu) & \alpha_M \\ \beta_M & -\beta_M \end{bmatrix}$$

After having defined these perturbed matrices, we have four variables at hand to change the coefficient set. Let us try it on the same example that we worked out in full detail. Our previous analysis is still valid up to the boundary equations. Since an analysis based on trying different values for all four parameters, would be too extensive and difficult to represent, we choose to vary only the parameters α_0 and α_M, the breakdown rates at two boundaries. These rates are anyway only meaningful relative to the breakdown rates at the boundaries. Thus, our analysis is actually complete in scope.

Using *Maple*, we are able to obtain the boundary equations as a function of the new parameters. We solve these equations and look at the proportion of the two coefficients as we already have done. By finding the proportion as both α_0 and α_M change in the interval $(10^{-10}, 10^{10})$ we arrive this three-dimensional graphical representation.

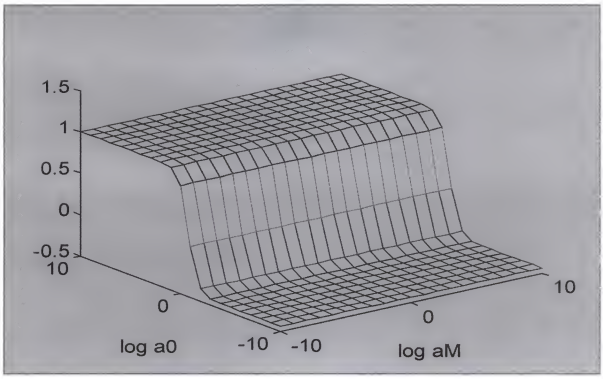

Figure 5.4 Proportion of coefficients, $\frac{w_{(1,1)}}{w_{(2,1)}}$, as a function of α_0 and α_M.

This graph clearly shows the validity of the bounds (5.15 and 5.16) we have derived. Moreover, one can conclude that the proportion tends asymptotically to these bounds as α_0 tends to zero and to infinity. An interesting fact is that it is not really dependent on α_M.

Furthermore, the proportion passes by zero around $\alpha_0 = 0.07$ (This figure is obtained by a bisection search). That means the effect of the first component can be fully eliminated when this boundary parameter is set accordingly. Under such conditions, the solution would be shaped only by the second component, except for the contribution

associated with the eigenvalue at the infinity. But this contribution would be limited to entries of the solution corresponding to full buffer.

Thus, it is indeed possible to change the systems behavior just by perturbing the boundary conditions. From an engineering perspective this result is quite important, since it gives us a tool for altering the systems behavior significantly by only adjusting the policies that relates to full and empty buffer processes.

This completes our analysis of the Model 2 (Exponential Server with Breakdown and Repair, Limited Buffer). We continue with the analysis of the following models. But having vividly illustrated the use of the spectral theory on this model, we shall not pursue the next models with an equal zeal.

Model 3: Server With Erlang-2 Service Times, Limited Buffer:

This model is quite close to the basic M/M/1. The only difference is that the processing time distribution, instead of being a mere exponential, is an Erlang-2 distribution, which is essentially the sum of the exponential distributions with identical parameters.

The matrix polynomial that corresponds to this QBDP is

$$L(x) = \begin{bmatrix} -(\lambda+\mu)x+\lambda & \mu x \\ \mu x^2 & -(\lambda+\mu)x+\lambda \end{bmatrix} \tag{5.17}$$

The characteristic equation can be easily derived from the matrix polynomial:

$$\det(L(x)) = -\mu^2 x^3 + (\lambda^2+\mu^2+2\lambda\mu)x - 2\lambda(\lambda+\mu)x + \lambda^2 \tag{5.18}$$

According to the theory, one would expect to have four eigenvalues corresponding to a QBDP with submatrices of dimension 2x2. Yet the characteristic equation being a

polynomial of degree three, can only provide us with three eigenvalues. Yet, since the nullspace of A is of dimension one, we have another eigenvalue at the infinity.

Another root that is always known is one by Property 4.3. This fact is used to check the validity of our derivation, and to ease the retrieval of the other two finite roots. Once the det(L(x)) of (5.18) is divided to (x-1), one obtains a second degree polynomial that embeds the critical two eigenvalues of the system:

$$\frac{\det(L(x))}{x-1} = -\mu^2 x^2 + \lambda(\lambda + 2\mu)x - \lambda^2 \tag{5.19}$$

The two eigenvalues embedded in the equation are critical since they are dependent on the system's parameters. Thereby, they elucidate the relation between the system parameters and the nature of the steady-state probabilities. In order to find and classify the nature of the eigenvalues one should first investigate the discriminant of the second-degree equation of (5.19).

$$\Delta = \lambda^2\left(\lambda^2 + 4\lambda\mu\right) \tag{5.20}$$

Since all the parameters are positive, one can conclude without much difficulty that the discriminant is always positive. This is important since it shows that the eigenvalues of the system are all real. That is no complex eigenvalue, which could cause an oscillating set of steady-state probabilities.

At this stage one can express in closed-form all the eigenvalues of the system:

$$x_1 = \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2\left(\left(1+2\left(\frac{\lambda}{\mu}\right)^{-1}\right) - \sqrt{1+4\left(\frac{\lambda}{\mu}\right)^{-1}}\right) \tag{5.21}$$

$$x_1 = \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2\left(\left(1+2\left(\frac{\lambda}{\mu}\right)^{-1}\right) + \sqrt{1+4\left(\frac{\lambda}{\mu}\right)^{-1}}\right) \tag{5.22}$$

$x_3 = 1 \quad , \; x_4 = \infty$

As one can see, the first two eigenvalues are more critical for the understanding of the system, since they depend on the system parameters. The eigenvalue at one will not affect the solution by Theorem 4.8. At this point, one needs to address issues like the dependence of two eigenvalues (5.21 and 5.22) to the system parameters and their range of values.

One thing that can be observed right away is that the square root of the discriminant (5.20) is always smaller than $\left(1+2\left(\frac{\lambda}{\mu}\right)^{-1}\right)$; thence, both eigenvalues (5.21 and 5.22) are always positive. If not, oscillatory behavior could have been observed in the solution.

One should notice at this stage, both model parameters have the unit, 1/[Time Unit]. Therefore their actual values depend on the choice of time unit. Yet one can express the eigenvalues in terms of parameters that are independent of the choice of time unit, while decreasing the number of parameters by one. This is done by the introduction of ρ, the traffic intensity, as we did in model one. Yet, the definition of the traffic intensity for our model would be a little bit different, $\rho = 2\frac{\lambda}{\mu}$. This is due to the fact that the processing time for the system is not $\frac{1}{\mu}$, but $\frac{2}{\mu}$.

When (5.21) and (5.22) are expressed in terms of the new parameters, one gets

$$x_1 = \frac{1}{8}\rho^2\left(\left(1+4\rho^{-1}\right) - \sqrt{1+8\rho^{-1}}\right) \tag{5.23}$$

$$x_1 = \frac{1}{8}\rho^2\left(\left(1+4\rho^{-1}\right)+\sqrt{1+8\rho^{-1}}\right) \tag{5.24}$$

At this stage, it is useful to investigate the conduct of the two eigenvalues as a function ρ. This would enables us to understand how the systems behavior changes as the traffic intensity changes.

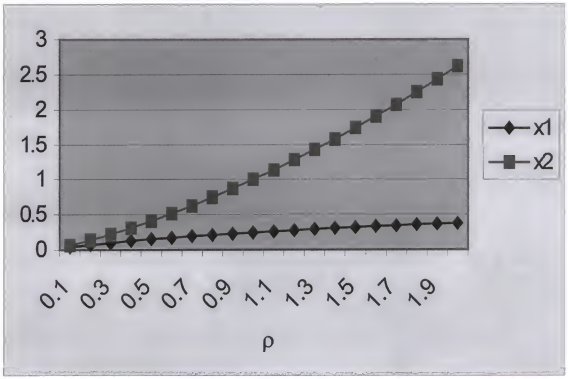

Figure 5.5 Eigenvalues drawn as a function of ρ.

Figure 5.5 is much more useful than the ones we drew for Model 2, since it illustrates all the values that the eigenvalues can take. This is due to the fact that the eigenvalues for the model can be expressed as a function of only one parameter, the traffic intensity.

Figure 5.5 clearly shows what happens as the traffic intensity changes. Both eigenvalues start at zero. As the traffic intensity increases, they both increase in a monotonic fashion. Yet while one always remains below the unity level, the other one transcends the unity level when the traffic intensity reaches one and then tends to infinity.

This fact is of utmost importance since it complies with our conjecture linking the physical balance to an eigenvalue at the unity. As we know, the system is in balance when the arrival rate is equal with the processing rate, that is, when the traffic intensity is one. The graph lucidly links the balance phenomenon to the eigenvalue at one. It is also interesting that the eigenvalue has the value one, only at one point.

Having observed the eigenvalues' conduct as the traffic intensity changes, we attempt to formalize our inferences by the use of mathematical techniques. We first set to investigate the limiting behavior of the eigenvalues–that means where they tend as the parameters take their limiting values, 0 and infinity.

As ρ tends to zero from the right, the both eigenvalues tend to zero. The first eigenvalue tends asymptotically to the function $\frac{2-\sqrt{2}}{4}\rho$, and the second one tends asymptotically to the function $\frac{2+\sqrt{2}}{4}\rho$.

As ρ tends to infinity, the second eigenvalue tends to positive infinity asymptotically to the function $\frac{1}{4}\rho^2$. The analysis of the first eigenvalue requires a successive use of L'Hopital's rule.

$$\lim_{\rho\to\infty} x_1 = \lim_{\rho\to\infty} \frac{1}{8}\frac{(1+4\rho^{-1})-\sqrt{1+8\rho^{-1}}}{\rho^{-2}} = \lim_{\rho\to\infty} \frac{1}{8}\frac{-4\rho^{-2}-\frac{1}{2}\left(1+8\rho^{-1}\right)^{-\frac{1}{2}}(-8\rho^{-2})}{-2\rho^{-3}}$$

$$= \lim_{\rho \to \infty} \frac{1}{4} \frac{1 - \left(1 + 8\rho^{-1}\right)^{\frac{1}{2}}}{\rho^{-1}} = \lim_{\rho \to \infty} \frac{1}{4} \frac{-\frac{1}{2}\left(1 + 8\rho^{-1}\right)^{\frac{3}{2}}(-8\rho^{-2})}{-\rho^{-2}} = 1$$

Thereby, the first eigenvalue tends to one as the traffic intensity tends to infinity. This indeed shows that the second eigenvalue never crosses the unity level.

At this stage, to show that the eigenvalues are monotonically increasing functions of the traffic rate, ρ, we examine the eigenvalues' derivatives with respect to ρ:

$$\frac{dx_1}{d\rho} = \frac{1}{4}\left(\rho + 2 - \left(\rho^4 + 8\rho^3\right)^{-\frac{1}{2}}(2\rho^3 + 6\rho^2)\right)$$

$$\frac{dx_2}{d\rho} = \frac{1}{4}\left(\rho + 2 + \left(\rho^4 + 8\rho^3\right)^{-\frac{1}{2}}(2\rho^3 + 6\rho^2)\right)$$

When the roots of both of the derivatives is investigated using a symbolic mathematical tool called *Maple*, one finds out both of them have no roots. Thence, both derivatives are always positive. This means that the derivatives are always strictly positive. Thus, both eigenvalues are monotonically increasing functions with respect to ρ.

At this point, one can conclude that the both of the components are more biased towards the smaller buffer levels when the traffic intensity is relatively small. As the traffic intensity increases, this bias becomes less and less evident. At one point, the second eigenvalue transcends the unity. After this point, the second component is biased towards the larger buffer sizes. Yet, the first component is always more prominent at smaller buffer levels.

The point at which the second eigenvalue crosses the unity is also of utmost importance. At that point the second eigenvalue is at unity, which means that the

component corresponding to it is balanced over the buffer levels. We have already postulated that an additional eigenvalue at one always relates to the balance of the system. Using *Maple* one can obtain the roots of the equation, $x_2 = 1$. The solution to this equation is $\rho = 1$. This result confirms our observation from the graph. Thus, the eigenvalue at one, insures a balanced system.

Having fully investigated the eigenvalues of the model, we complete the analysis of Model 2 using the spectral theory. One should notice that the organization of the section follows the previous one closely. This is due to the fact that the methodology applied for Model 1, is readily applicable to all other models.

Model 4: Server With Erlang-2 Service Times, Breakdown and Repair, Limited Buffer:

This model incorporates elements of both Model 2 and Model 3. There is a breakdown repair mechanism as in Model 2, and the production times are distributed with an Erlang-2 distribution as in Model 3. Using the derived submatrices from Chapter 3, one can write the matrix polynomial that corresponds to this QBDP:

$$L(x) = \begin{bmatrix} -(\alpha+\lambda+\mu)x+\lambda & \mu x & \alpha x \\ \mu x^2 & -(\alpha+\lambda+\mu)x+\lambda & \alpha x \\ \beta x & 0 & -(\beta+\lambda)x+\lambda \end{bmatrix} \tag{5.25}$$

The characteristic equation can be easily derived from the matrix polynomial:

$$\begin{aligned} \det(L(x)) = {} & \mu^2(\beta+\lambda)x^4 - (\alpha^2\lambda+\lambda^3+\beta\lambda^2+\alpha\beta\lambda+2\alpha\lambda^2+\beta\mu^2+2\beta\lambda\mu \\ & +2\lambda^2\mu+2\alpha\lambda\mu+2\lambda\mu^2)x^3+\lambda(\alpha^2+4\alpha\lambda+\alpha\beta+2\beta\mu+4\lambda\mu \\ & +2\beta\lambda+2\alpha\mu+3\lambda^2+\mu^2)x^2-\lambda^2(\beta+3\lambda+2\mu+2\alpha)x+\lambda^3 \end{aligned} \tag{5.26}$$

According to the theory, one would expect to have six eigenvalues corresponding to a QBDP with submatrices of dimension 2x2. Yet, the characteristic equation being a

polynomial of degree four can only provide us with four eigenvalues. But, since the nullspace of A is of dimension two, we have two other eigenvalues at the infinity.

Another root that is always known is one by Property 4.3. This fact is used to check the validity of our derivation, and to ease the retrieval of the other three finite roots. Once the det(L(x)) of (5.26) is divided to (x-1), one obtains a second degree polynomial that embeds the three critical eigenvalues of the system:

$$\frac{\det(L(x))}{x-1} = \mu^2(\lambda+\beta)x^3 - \lambda(2\alpha\lambda + 2\lambda\mu + \alpha^2 + \mu^2 + \lambda^2 + \beta\lambda + \alpha\beta + 2\alpha\mu + 2\beta\mu)x^2 + \lambda^2(2\mu + \beta + 2\lambda + 2\alpha)x - \lambda^3 \quad (5.27)$$

The two eigenvalues embedded in the equation are critical since they are dependent on the system's parameters. Thereby, they elucidate the relation between the system parameters and the nature of the steady-state probabilities. Yet, since the equation we have on hand is not as simple as the ones we had previously, we had to rely on a symbolic mathematics tool, *Maple*. Although we were able to obtain the three roots, their closed form expressions were complicated to an extent that just to cite them here would require more than a page.

Under these circumstances, one cannot delve into the analysis of the eigenvalues as they relate to system parameters. Thence, only a numerical analysis is plausible. For such an endeavor, one would need to deal with four parameters, as it is the case with Model 3. Yet with the introduction of parameters that are not dependent on choice of time unit, the number of parameters that need to be considered would be reduced to three. These parameters are almost the same as the ones in Model 2:

$\rho = 2\frac{\lambda}{\mu}$, Traffic intensity.

$\gamma = \frac{\alpha}{\lambda}$, Breakdown intensity.

$\delta = \frac{\beta}{\lambda}$, Repair intensity.

Thus, for Model 4 we are unable to get closed form solutions. Without the guidance of the closed-form expression a numerical study needs to be quite extensive. We relegate such numerical analysis of this model to future research. We still find it appropriate to give a graphical representation–for an instance that has been investigated for Model 2–for the behavior of the three finite eigenvalues as the traffic intensity changes in Figure 5.6.

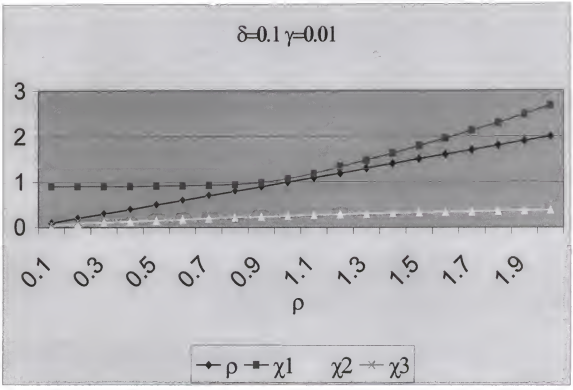

Figure 5.6 Three eigenvalues drawn as a function of ρ.

One can observe that the graph obtained is a mixture of the ones for Model 2 and Model 3. The first eigenvalue behaves just like the first eigenvalue of Model 2. The second eigenvalue is very much comparable to the second eigenvalue of Model 2. Finally, the third eigenvalue's conduct is like the one of the second eigenvalue of Model 3.

The similarities are to such an extent that we can postulate the theoretical results derived for Models 2 and 3, are applicable to this model, although the interaction between the elements of this model makes a similar derivation impossible. We can even observe that, the first eigenvalue is the only eigenvalue that crosses the unity level as it is with previous models. Furthermore, the crossing occurs at $\rho = \frac{\delta}{\delta + \gamma}$, which is the point that the system is balanced.

Since Model 4 is a hybrid of the two aforementioned models, one sees the strong coupling between the model characteristics and the behavior of eigenvalues. This phenomenon certainly supports our hindsight that the eigenvalues are critical in the characterization of the models.

Model 5: Two Exponential Servers With One Intermediate Limited Buffer

In this model, two exponential servers are present and the buffer is placed between the servers. The typical two-server model in the literature assumes an infinite source for the server, which is stochastically the equivalent of Model 1. The input is a Poisson process. One can actually assume that there is a second buffer at the input of the

system. But the size of this buffer would be limited to one. Therefore, this model is a good starting point for the examination of the production lines with two buffers.

The matrix polynomial that corresponds to this QBDP is

$$L(x) = \begin{bmatrix} \mu_2 x^2 - (\lambda + \mu_2)x & \lambda x \\ \mu_1 & \mu_2 x^2 - (\mu_1 + \mu_2)x \end{bmatrix} \tag{5.28}$$

The characteristic equation can be easily derived from the matrix polynomial

$$\begin{aligned} \det(L(x)) = {} & \mu_2{}^2 x^4 - \mu_2(\mu_1 + 2\mu_2 + \lambda)x^3 \\ & + (\lambda\mu_1 + \lambda\mu_2 + \mu_1\mu_2 + \mu_2{}^2)x^2 - \lambda\mu_1 x \end{aligned} \tag{5.29}$$

According to the theory, one would expect to have four eigenvalues corresponding to a QBDP with submatrices of dimension 2x2. We can readily see that one of the eigenvalues is zero, since there is no constant term in (5.29). One could have predicted this root at zero by observing the fact that the nullspace of C is of dimension one.

We have no eigenvalue at infinity for this model. Since this eigenvalue relates directly to the zero buffer level, there will be a special term in the solution that only affects the zero buffer level probabilities. It is possible to link this to the possible starvation phenomenon in this model. In the previous models at which there was no server in front of the buffer, there was an eigenvalue at the infinity, which can be linked to the blocking of the arrivals once the buffer is full.

Another root that is always known is one by Property 4.3. This fact is used to check the validity of our derivation, and to ease the retrieval of the other two finite roots. Once the det(L(x)) of (5.29) is divided to x.(x-1), one obtains a second degree polynomial that embeds the critical two eigenvalues of the system:

$$\frac{\det(L(x))}{x.(x-1)} = \mu_2{}^2 x^2 - \mu_2(\mu_1 + \mu_2 + \lambda)x + \lambda\mu_1 \tag{5.30}$$

The two eigenvalues embedded in the equation are critical since they are dependent on the system's parameters. Thereby, they elucidate the relation between the system parameters and the nature of the steady-state probabilities. In order to find and classify the nature of the eigenvalues one should first investigate the discriminant of the second-degree equation of (5.30).

$$\Delta = \mu_2{}^2\left((\mu_1 + \mu_2 + \lambda)^2 - 4\lambda\mu_1\right) \tag{5.31}$$

It is possible to write the discriminant in another form:

$$\Delta = \mu_2{}^2\left((\mu_1 + \mu_2 - \lambda)^2 + 4\lambda\mu_2\right) \tag{5.32}$$

This form is useful since it shows that the discriminant is always positive. Since it is known the parameters are always positive, the expression $\mu_2{}^2\left((\mu_1 + \mu_2 - \lambda)^2 + 4\lambda\mu_2\right)$ is always positive. Thereby, the discriminant is always positive. This is important since it shows that the eigenvalues of the system are all real. That is, no complex eigenvalue, which could cause an oscillating set of steady-state probabilities, exists.

At this stage one can express in closed-form all the eigenvalues of the system:

$$x_1 = \frac{1}{2}\left(\left(1 + \frac{\mu_1}{\mu_2} + \frac{\lambda}{\mu_2}\right) - \sqrt{\left(1 + \frac{\mu_1}{\mu_2} + \frac{\lambda}{\mu_2}\right)^2 - 4\frac{\lambda}{\mu_2}\frac{\mu_1}{\mu_2}}\right) \tag{5.33}$$

$$x_1 = \frac{1}{2}\left(\left(1 + \frac{\mu_1}{\mu_2} + \frac{\lambda}{\mu_2}\right) + \sqrt{\left(1 + \frac{\mu_1}{\mu_2} + \frac{\lambda}{\mu_2}\right)^2 - 4\frac{\lambda}{\mu_2}\frac{\mu_1}{\mu_2}}\right) \tag{5.34}$$

$x_3 = 0 \quad , x_4 = 1$

As one can see, the first two eigenvalues are more critical for the understanding of the system, since they depend on the system parameters. The eigenvalue at one will not affect the solution by Theorem 4.8. At this point, one needs to address issues such as the dependence of two eigenvalues (5.13 and 5.14) to the system parameters and their range of values.

One thing that can be observed right away is that the square root of the discriminant (5.31) is always smaller than $\left(1+\frac{\mu_1}{\mu_2}+\frac{\lambda}{\mu_2}\right)$; thence, both eigenvalues (5.33 and 5.34) are always positive. If not, oscillatory behavior could have been observed in the solution.

All the parameters that we make use of here have the unit, 1/[Time Unit]. Therefore their actual values depend on the choice of time unit. Yet one can express the eigenvalues in terms of parameters that are independent of the choice of time unit, while decreasing the number of parameters by one. For this model, we need to define two parameters:

$\rho = \frac{\lambda}{\mu_2}$, Traffic intensity.

$\nu = \frac{\mu_1}{\mu_2}$, First stage intensity.

These definitions are useful from an engineering point of view, since it relates the eigenvalues of the system, to physically meaningful quantities. Thence, they make the relation between the system assumptions and the eigenvalues, which determine the conduct of the system, more visible.

When (5.33) and (5.34) are expressed in terms of the new parameters, one gets

$$x_1 = \frac{1}{2}\left((1+\upsilon+\rho) - \sqrt{(1+\upsilon+\rho)^2 - 4\upsilon\rho}\right) \tag{5.35}$$

$$x_1 = \frac{1}{2}\left((1+\upsilon+\rho) + \sqrt{(1+\upsilon+\rho)^2 - 4\upsilon\rho}\right) \tag{5.36}$$

At this stage, it is useful to draw these two eigenvalues as a function ρ, for certain physically meaningful choice of parameters γ and δ. This would enables us to understand how the systems behavior changes as the traffic intensity changes.

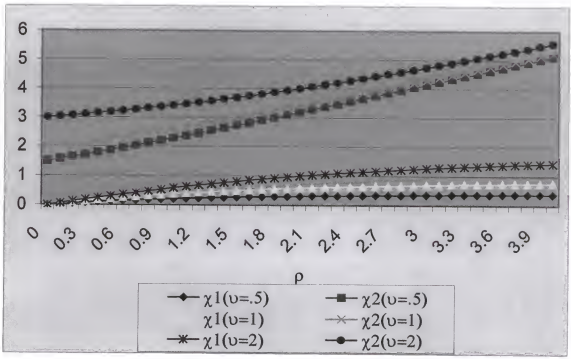

Figure 5.2 Eigenvalues drawn as a function of ρ, functions for different system parameters, υ.

Figure 5.1 clearly shows what happens as the traffic intensity changes. The first eigenvalue starts above the unity level and increases in a linear fashion as ρ tends to infinity. The second eigenvalue starts at zero and then asymptotically approaches an

upper bound. It may or it may not cross the unity level, depending on the choice of first stage intensity parameter. Thus, in order that a balance point exist for this model, this parameter needs to be higher than a certain level.

At this stage, we attempt to formalize our inferences by the use of mathematical techniques. We first set to investigate the limiting behavior of the eigenvalues–that means where they tend as the parameters take its limiting values, 0 and infinity.

$$\lim_{\rho \to 0^+} x_1 = 0$$

$$\lim_{\rho \to 0^+} x_2 = 1 + \upsilon$$

We observe that the second eigenvalues are above the unity level when the traffic intensity is zero. As ρ tends to infinity, $x_1 \to 0-0$, which is an indeterminate form. One can resort to L'Hopital's rule to overcome this and obtain

$$\lim_{\rho \to \infty} x_1 = \upsilon$$

This means, balance can occur in this system if and only if the first stage intensity is higher than unity. In other words, the processing rate of the first server should be higher than the one of the second server for this to happen. This occurrence can be attributed to the starvation phenomenon.

As ρ tends to infinity, $x_2 \to \rho$. Thus, the second eigenvalue asymptotically approaches to the linear function $f(\rho) = \rho$.

We can readily obtain the results for υ, because of the interchangeability of the variables in the equations. That means

$$\lim_{\upsilon \to 0^+} x_1 = 0$$

$$\lim_{\upsilon \to 0^+} x_2 = 1 + \rho$$

$$\lim_{\upsilon \to \infty} x_1 = \upsilon$$

$$\lim_{\upsilon \to \infty} x_2 = \infty$$

We prove that the eigenvalues are monotonically increasing functions of ρ, by taking their derivatives and showing that the derivatives do not have roots using *Maple*. We also find the point where the first eigenvalue crosses unity level, the balance point, by using *Maple*. The balance point is

$$\rho = \frac{\upsilon}{\upsilon - 1} \tag{5.37}$$

As one can observe, υ needs to be greater than one for the balance to exist. Furthermore, as first buffer intensity increases, the system reaches balance for smaller values of traffic intensity. As υ tends to infinity, the balance point tends to one; as υ tends to one from the right, the balance points tends to infinity.

At this stage it is possible to interpret the physical meaning of the balance point. One can check that the balance point is the point at which the average traffic at downstream line would be equal to the average traffic at upstream line if the system was to be divided to two by taking out the buffer corresponding to the counting set. For production lines, the average traffic that goes through the system is the production rate of the line.

If one takes out the buffer of Model 5 out and, thereby, eradicates all possibility of blocking for the upstream line, the production rate of the system can be found with a

simple Markov analysis. This rate is $\frac{\mu_1 \lambda}{\mu_1 + \lambda}$. On the other hand, if the downstream line is fed by an infinite supply of parts, the production rate of the downstream line is the stand-alone production rate of the second station, μ_2. The balance with respect to the buffer occurs when these two rates are equal:

$$\mu_2 = \frac{\mu_1 \lambda}{\mu_1 + \lambda} \tag{5.38}$$

By simple algebraic manipulation, it is possible to show that equations (5.37) and (5.38) are equivalent. Thus, we conjure that when the described balance occurs around the buffer corresponding to the counting set of the QBDP, an additional eigenvalue is at the unity. Thus, the balance of a QBDP corresponds to such a physical balance phenomenon for models of production systems.

This completes our analysis of the Model 5 (Two Exponential Servers With One Intermediate Limited Buffer). Unfortunately, a theoretical analysis of the Models 6,7, and 8 is not possible because of their large size. They result in closed-form expressions that are so extensive that it is not even possible to print them in this work. Moreover, the size of submatrices is not defined for Models 7 and 8. Thus, they require an extensive analysis elucidating the effect of the submatrix size.

We find it appropriate not to incorporate their analyses, which are rather numerically geared. Thus, we complete the application of spectral methods for the analysis of production line models at this point.

5.4 Common Spectral Characteristics of Models of Production Lines

In the previous section, we have demonstrated how the spectral theory we have derived can be applied to models of production lines. For this purpose we used the QBDP formulations of specific production lines, which we derived in Section 3.3. As a result of this effort, we were able to discern many common traits in the models. Although this section, which tries to elucidate the common characteristics of production lines, is based only on the five models we analyzed, the cogency of our observations suggests that our conclusions are quite general. They may even be applicable to QBDP models that pertain to other areas. Yet, one would need to interpret them according to their specific context.

Our first observation is the fact that all eigenvalues that pertain to models of production lines are positive and non-complex. This should be due to the special characteristics of the submatrices that are involved in QBDPs. It may be possible to prove this by using the properties of stochastic submatrices. This may be a worthwhile addition to our work.

An immediate consequence of the positive, non-complex nature of the eigenvalues is the fact that the *components* do not exhibit oscillatory behavior. That means, a *component*'s effect is either monotonically increasing or monotonically decreasing as the higher buffer levels are considered, or it is constant for all buffer levels. That means a *component* is either biased towards lower buffer levels or higher buffer levels.

Another interesting observation is that the eigenvalue functions never intersect each other, with the exception of the eigenvalue at the unity being intersected only once by only one of the eigenvalue functions. That means multiplicity only occurs at the unity, not for any other value.

Furthermore, the aforementioned intersection is of considerable interest, since it occurs when the system is physically balanced around the buffer. Here, we define a new balance concept with a buffer-centric approach. The system is balanced with respect to a buffer–according to our definition, –when the production rate of the line that is downstream with respect to the buffer is equal to the production rate of the line that is upstream with respect to the buffer, given that the buffer is taken out of the system and the system is divided to two parts.

Using the fact that the boundary equations do not effect the location of the balance point, we can conclude that the balance of a system that can be modeled by a QBDP is independent of the boundary behavior of the system. This idea can be used in finding the balance point for quite complicated systems with the help of numerical analysis tools.

The contribution of this different balance concept is important, because it describes a balanced probability distribution for a production line, whereas the old balance concept, which relates to the equality of stand-alone production rates of all servers, cannot guarantee such a distribution. This balance definition takes into consideration the starvation and the blocking in the system, which truly affects the balance in the system.

Another observation is that, it is always possible to define traffic intensity for any production line. As the line gets more complicated, one needs to introduce some other

parameters. Yet, it is always of interest to look at the conduct of the eigenvalues as a function of the traffic intensity, since it provides a tool we can employ to compare different systems.

The eigenvalues at zero and infinity can be considered as correction masses that are employed for zero and full buffer levels. They strictly contribute to the steady-state probabilities at these points. They can be thus related to the phenomena such as starvation and blockage.

The knowledge of eigenvalues conduct can be very useful for making sensitivity analysis. One can check how much the systems behavior would change if the parameters were perturbed, by considering the locus of eigenvalues. This can be quite useful at the design of a production line.

Another fact that can be extremely useful, is the fact that the boundary equations can be used to manipulate the composition of final steady-state probabilities. We have already explained how each *component* is a solution candidate and the boundary equations determine how much of each component goes into the final result.

Yet, our experimentation with the models suggests that there is a strong coupling between the components. Even by applying great perturbations to the boundary equations, the amount of change in the composition of the final result is limited. But this can still be used to push the system in the desired direction to a certain extent. Moreover, this fact insinuates that the behavior of the system in the boundary is more critical than one may expect.

In retrospect, the spectral theory provided new tools that furthered our understanding of the behavior of the steady-state probabilities for models of production

lines. These tools are beneficial for the sensitivity analysis of the models. They can be quite instrumental at the design of the production lines and at pushing them towards a more desirable conduct.

CHAPTER 6
CONCLUSION

The work presented within this body is part of a much larger research attempt. At the onset of our attempt, we had in mind many objectives. The main thrusts of the research can be summarized as

1. Extension of the matrix polynomial method to production systems with more than one buffer,

2. Formalization of the matrix polynomial formulation for quasi-birth-processes,

3. Establishment of the relation between the theoretical properties and the properties of modeled systems, with emphasis on the analysis of the multiplicity cases, which arise as a natural phenomenon in balanced lines,

4. Establishment of the relation between the model assumptions and observed line properties for the production lines,

5. Analysis of the domain of eigenvectors.

We examined for a long time ways of solving the first problem, which had been attempted by many scholars before. Although we were able to discover a certain structure that was tantalizing, we were unable to exploit this structure in developing a solution procedure. Yet, we discovered some fractal-like structure in the transition rate matrix for the processes to model production systems with more than one buffer. We demonstrated this structure in Model 8 of Chapter 3. We also made a note that it can be actually

utilized in a procedure to rapidly create the transition rate matrices for such processes. After many attempts, we decided not to pursue this problem any further, since it was inherently difficult, even if possibly solvable with such techniques. Thus, we relegated this endeavor to future researchers.

For the second thrust of the research, we reached a successful culmination. The formal platform we have presented within the body of this work is the result of this thrust. This formal exposition was only possible with a thorough analysis that took into account multiplicity cases and with the introduction of generalized eigenvectors. We were able to unify all the previously known properties and the new properties, which we have stated, within this formal platform. By doing so, we proved that all *components*, which are constructed by an eigenvalue-eigenvector pair, show all the characteristics of the full solution. This is the main contribution in this research.

We also put thought in relating the theoretical properties we had derived to modeled systems properties. Particularly, we tried to give a physical interpretation to the phenomena that occurred in the solution procedure such as multiplicity. We observed that the eigenvalue at one, relates to the balance of the systems. If the system's feeding and output retrieval rates are the same, the multiplicity of the eigenvalue at one is at least two. Yet, these relations were not obvious and needed attentive examination to be fully exposed.

Thence, in order to elucidate these relations we analyzed a full line of production line models. We first derived the QBDP formulation for eight fundamentally significant production line models. We ipso facto demonstrated the strength of QBDP platform in modeling and found the spectral properties of these well-investigated traditional models.

Then, we employed these models as a test bed to demonstrate how the spectral theory we have derived for the QBDPs could be applied to given models. By doing so, we also elucidated, in an empirical fashion, how the eigenvalues of the system affect the final solution, the steady-state probabilities. Thus, we established the relation between the theoretical properties and the properties of modeled systems that was stated as the third thrust of our research.

We demonstrated the importance of the *components* we defined by showing how they shape the final solution. This led to the idea of making use of the eigenvalues as a design tool, since one could actually determine the systems behavior by looking at its eigenvalues. This is very similar to the crucial importance of zeros and poles of a linear time independent system in control theory. Having observed the importance of eigenvalues, we proposed using graphical tools that link the location of eigenvalues to the system parameters, in the spirit of zero, pole graphs of control methodology. Due to the non-complex nature of the eigenvalues we observed, this was possible with simpler function graphs.

All our analysis was on to the production systems, since the production systems were to constitute the concrete platform we elected for our investigation. Although the results obtained can always be extended to a number of other systems that can be modeled with QBDPs, the production systems, which had been extensively studied, are a critical point in case. By studying their spectral properties, and demonstrating the use of the results obtained for sensitivity analysis and design purposes, we made a noteworthy contribution to this line of work. Furthermore, we linked the theory we have presented to

the crux of the industrial research concerns, which is the design and the analysis of production systems.

For demonstrating the use of eigenvalues as a design and sensitivity analysis tool, we investigated models of production lines. We showed how the system parameters affected the *components*, and thereby the solution. Ipso facto, we established the relation between the model assumptions and observed line properties within the study. That was the fourth thrust that we stated.

We also observed the domain of eigenvalues in the matrix polynomial solution procedure for the quasi-birth-death-processes. We noticed that the roots of the characteristic equation for this problem were never complex or negative for the physically meaningful production line models we modeled. Although we were unable to prove this phenomenon formally, we demonstrated it on a variety of production line models we considered.

In retrospect, we developed a theory on the spectral properties of the QBDPs. Then, we evinced its use by applying it on a class of important models. Yet, there are areas that are left for future research:

1. An extension of the matrix polynomial method to QBDPs with more than one counting set, such as production systems with more than one buffer,

2. Using the additional structure of QBDP, proving the characteristics of the eigenvalues obtained using the matrix polynomial method (i.e. real, positive),

3. Application of the same methodology to other areas where QBDP arise (i.e. analysis of telecommunications systems).

Our hunch is that, the structure apparent in the QBDPs is yet to be fully exploited. That means it is an area full of prospects for future researchers. Our study will form a firm foundation for this direction of research.

LIST OF REFERENCES

Avi-Itzak, B. and Yadin, M. (1965), "A Sequence of Two Servers with No Intermediate Queue," Management Science, Vol. 11, No. 5, pp. 563-564.

Berman, O. (1982), "Efficiency and Production Rate of A Transfer Line with Two Machines and Finite Storage Buffer," European Journal of Operational Research, Vol. 9, pp. 295-308.

Buzacott, J.A. (1967), "Automatic Transfer Lines with Buffer Stocks," International Journal of Production Research, Vol. 5, No. 3, pp. 183-200.

Buzacott, J.A. and Kostelski, D. (1987), "Matrix-geometric and Recursive Algorithm Solution of A Two-stage Unreliable Production Line," IIE Transactions, Vol. 19, No. 4, pp. 429-438.

Dallery, Y., and Gershwin, S.B. (1992), "Manufacturing Flow Line Systems: A Review of Models and Analytical Results," Queuing Systems, Vol. 12, No. 3, pp. 3-94.

Evans, R.V. (1967), "Geometric Distribution in Some Two-dimensional Queuing Systems," Operations Research, Vol. 15, pp. 830-846.

Gershwin, S.B. and Berman, O. (1981), "Analysis of Transfer Lines Consisting of Two Unreliable Machines with Random Processing Times and Limited Storage Buffers," AIIE Transactions, Vol. 13, No. 1, pp. 2-11.

Hajek, B. (1982), "Birth and Death Processes on the Integers with Phases and General Boundaries," Journal of Applied Probability, Vol. 19, No. 3, p.p. 488-499.

Hatcher, J.M. (1969), "The Effect of Internal Storage on the Production Rate of a Series of Stages Having Exponential Service Times," AIIE Transactions, Vol. 1, No. 2, pp. 150-156.

Hillier, F.S. and Boiling, R.W. (1967), "The Effect of Some Design Factors on the Efficiency of Production Lines with Variable Operations Times," Journal of Industrial Engineering, Vol. 17, No. 12, pp. 651-658.

Hunt, G.C. (1956), "Sequential Arrays of Waiting Lines," Operations Research, Vol. 4, No. 6, pp. 674-684.

Kleinrock, L. (1975), Queuing Systems, Volume 1: Theory, John Wiley & Sons Inc., New York.

Koenigsberg, E. (1959), "Production Lines and Internal Storage–A Review," Management Science, Vol. 5, pp. 410-433.

Lancester, P., Gohlberg, I. and Rodman, L. (1982), Matrix Polynomials, Academic Press Inc., New York.

Muth, E.J. (1979a), "The Reversibility Property of Production Lines," Management Science, Vol. 25, No. 2, pp. 152-158.

Muth, E.J. (1979b), "Models of Production Lines in Which Stations Are Subject to Breakdown," Methods of Operations Research, Vol. 35, pp. 345-346.

Neuts, M.F. (1981), Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach, Johns Hopkins Press, Baltimore.

Okumura, K. and Yamashina, H. (1977), "Analysis of Effect of Buffer Storage Capacity in Transfer Line Systems," AIIE Transactions, Vol. 9, No. 2, pp. 127-135.

Rao, N.P. (1975), "Two Stage Production System with Intermediate Storage," AIIE Transactions, Vol. 7, No. 4, pp. 414-421.

Sastry, B.L.N. (1985), "Analysis of Two-Machine Markovian Production Lines and Decomposition of Longer Lines," Ph.D. Thesis, Industrial Engineering and Operations Research Group, Indian Institute of Technology Bombay.

Sevast'Yanov, B.A. (1962), "Influence of Stage Bin Capacity on the Average Standstill Time of a Production Line," Theory of Probability and Its Applications, Vol. 7, No. 4, pp. 429-438.

Sheskin, T. J. (1974), "Allocation of Interstage Storage Along an Automatic Production Line," AIIE Transactions, Vol. 8, No. 1, pp. 146-152.

Shick, I.C. and Gershwin, S.B. (1978), "Modeling and Analysis of Unreliable Transfer Lines with Finite Interstage Buffers," Report ESL-FR-834-6, Massachussets Institute of Technology, Cambridge.

Tan, B. (1994), "A Decomposition Method for Multistation Production Systems", Ph.D. Dissertation, Industrial and Systems Engineering, University of Florida, Gainesville.

Tan, B. and Yeralan, S. (1997) "Analysis of Multistation Production Systems with Limited Buffer Capacity Part 2: The Decomposition Method," Mathl. Comput. Modelling, Vol. 25, No. 11, pp. 109-123.

Vladzievki, A.P. (1951), "Probabilistic Law of Operation and Internal Storage of Production Lines," Automatika I Telemkhana, Vol. 13, No. 3, pp. 227-281 (in Russian).

Wallace, V. (1969), "The Solution of Quasi Birth and Death Processes Arising from Multiple Access Computer Systems," Ph.D. Thesis, Systems Engineering Laboratory, University of Michigan, Ann Arbor.

Yeralan, S. and Muth, E.J. (1987), "A General Model of a Production Line with Intermediate Buffer and Station Breakdown," IIE Transactions, Vol. 19, No. 2, pp. 130-139.

Yeralan, S. and Tan, B. (1997), "Analysis of Multistation Production Systems with Limited Buffer Capacity Part 1: The Subsystem Model," Mathl. Comput. Modelling, Vol. 25, No. 7, pp.109-122.

BIOGRAPHICAL SKETCH

Murat Fadiloglu was born on May 12, 1970, in Izmir, Turkey. He obtained a bachelor's degree in electrical engineering with a concentration in control and telecommunications from Bilkent University in Ankara, Turkey, in June 1993. He continued his studies at the University of Florida, where he earned a master's degree in industrial and systems engineering, in August 1995. During his time at the University of Florida, he worked as a teaching assistant in the Department of Industrial and Systems Engineering, and as a research assistant in the Center for Intelligent Machines and Robotics in the Department of Mechanical Engineering. He is currently an instructor in the Department of Industrial and Systems Engineering.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sencer Yeralan, Chairman
Associate Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Eginhard J. Muth
Emeritus Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Suleyman Tufekci
Associate Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sherman X. Bai
Assistant Professor of Industrial and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John K. Schueller
Associate Professor of Mechanical Engineering

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1998

Winfred M. Phillips
Dean, College of Engineering

Karen A. Holbrook
Dean, Graduate School